Essays on Applied Economics

Fatemeh Saberianranjbar

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree: Doctorate in Philosophy Economics

Department of Economics
Faculty of Social Sciences
University of Ottawa

Abstract

Chapter 1.—In the 1970s, competition policy in the United States banking sector changed from exempting competition to liberalization and deregulation. Competition not only plays an important role in allocational efficiency but it is also essential for long-term economic growth. This chapter develops a model of banking contributions to evaluate to what extent banks affect the level of competition in the banking sector, and tests the model's predictions using a novel detailed dataset which includes all contributions made by banks from 1993 to 2010 in the United States. Controlling for banks' characteristics, the results are consistent with the model's predictions and show show that a higher level of contributions increases the Lerner index (as a measure of competition) or in the other words, decreases the level of competition.

Chapter 2.—This chapter provides the first empirical evidence that market structure affects the electoral power of firms as special interest groups. Firms not only affect the election outcomes by making contributions to their preferred candidates, they also enforce social norms among their members by encouraging them to vote for the candidate with the most closely-aligned interests. This chapter uses a linear probability model to analyze 574 open-seat races for the House of Representatives in the United States between 1990 and 2014. The results show that, even when controlling for the total value of contributions made to a candidate, political donations made by firms with high market power have a positive effect on the candidates' probability of winning. The findings are consistent with the idea from collective action theory that concentrated industries are more likely to behave as an

organized interest group to advance their interests.

Chapter 3.—Ethnic heterogeneity is an important factor in the formation of human sexual network and the prevalence of STDs. Racial and ethnic ties create closed social networks with rigid in-group boundaries and hampers the intra-group dissemination of information. Slow information flow among groups facilitates the spread of STDs by encouraging individuals to ethnically diversify their sexual partners in order to lower the chance of getting caught cheating. Analyzing a cross-province sample of 39,830 sexually active adults driven from the 2013-2014 Canadian Community Health Survey, we find that individuals who live in a highly ethnically diversified neighbourhood are more susceptible to STDs compared to ones who live in a ethnically homogenous neighbourhood. Evidence from several robustness checks suggests that the relationship is causal.

Declaration

All chapters of this thesis are self-containing research articles. Chapters 1 and 3 are from joint research. The first chapter is co-authored with Aggey Semenov and Hećtor Peŕez Saiz. I acknowledge the contribution of Roland Pongou for the research associated with the third chapter of this thesis. For this chapter, his contribution is equal to my own.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisors Dr. Aggey Semenov and Dr. Hećtor Peŕez Saiz for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. My sincere thanks also goes to Dr. Roland Pongou who provided me an opportunity to join his team, and who has been a tremendous mentor for me. Without his precious support it would have not been possible to finish this thesis.

Besides my advisers, I would like to thank the rest of my thesis committee: Dr. Jean-Thomas Bernard, Dr. Pierre Brochu, Dr. Christopher M. Gunn and Dr. Matthieu Chemin for their time, interest, and insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives. I would like to acknowledge Dr. Catherine Deri Armstrong, Dr. Jason Garred for their support and advices. A special thanks to Maxime Dufournaud-Labelle for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years.

Last but not the least, I would like to thank my family for all their love and encouragement. My parents who allowed me to realize my own potential and have cherished with me every great moment and supported me whenever I needed it. I cannot thank my loving and supportive sister, Soodeh, enough for encouraging me throughout this experience.

Contents

Abstract ii

Declaration iv

Acknowledgments v

General Introduction 1

1 The Effect of Campaign Contributions on Competition in the U.S. Banking Sector 5

1.1 Introduction 5
1.2 Literature review 7
1.3 Banks lobbying and the level of competition 10
1.3.1 Model 10
1.3.2 The equilibrium 13
1.3.3 Government program 14
1.3.4 The lobbying game 15
1.3.5 Testable implication 17
1.4 Data 17
1.5 Empirical analysis 20
1.5.1 The Lerner index as a measure of competitiveness 20
1.5.2 The effect of campaign contributions on the Lerner index 23
1.6 Results 24
1.7 Robustness checks 25
1.8 Conclusion 27

2 Market Structure and the Electoral Power of Special Interest Groups 44

2.1 Introduction 44
2.2 Literature review 46

2.3 The mechanism of influence . . . 48
2.3.1 Testable hypotheses . . . 51
2.4 Data . . . 51
2.4.1 Campaign contributions data . . . 52
2.4.2 Concentration ratio data . . . 53
2.4.3 Election results data . . . 54
2.5 Methodology . . . 57
2.6 Results . . . 59
2.6.1 Main results . . . 59
2.6.2 Assessing the relative contribution of the components of the contributions on the election outcomes . . . 60
2.6.3 Robustness . . . 61
2.6.4 Alternative standard errors . . . 63
2.7 Conclusion . . . 63

3 Ethnic Heterogeneity and STDs in Canada† 79
3.1 Introduction . . . 79
3.2 Literature review . . . 82
3.3 Mechanism of influence . . . 84
3.3.1 Testable implications . . . 86
3.4 Ethnic diversity . . . 87
3.5 Data and empirical strategy . . . 87
3.5.1 Data . . . 87
3.5.2 Identification strategies . . . 91
3.6 Results . . . 96
3.6.1 Main . . . 96
3.6.2 Instrumental variables . . . 99
3.6.3 Falsification test: The effect of ethnic diversity of the non-transmissible diseases . . . 100
3.7 Conclusion . . . 101

Bibliography 117

List of Figures

1.1 The trend of contributions by bank from 1993 to 2010 19

2.1 The distribution of the vote share of the Democrat and Republican candidates in open-seated races from 1990 to 2014 56
2.2 Regression discontinuity design of the open-race election of the U.S. House of Representative between 1990-2014 67
2.3 Total amount received by candidates (1990-2014) 70
2.4 Closeness of the races 73
2.5 The Relation between concentration ratio and the average size of firms operating in that industry 78

3.1 Forward Sortation Area* 102
3.2 Ethnic diversity* 103
3.3 Distribution of ethnic diversity index (HHI)* 104
3.4 Effect of FSA-level ethnic heterogeneity on the number of sexual partners* . 105
3.5 Effect of FSA-level ethnic heterogeneity on the prevalence of STDs* 106

List of Tables

1.1 Summary statistics from 1993 to 2010 29
1.2 Estimation of total cost of banks 30
1.3 The effect of contributions on the Lerner index 38
1.4 The effect of contributions on the Lerner index (2 lags) 39
1.5 The effect of contributions on H-Statistics 40
1.6 Estimation of total revenue of banks 41

2.1 Summary statistics of the candidates' vote share in open-seated races between1990 and 2014 56
2.2 The effect of party winning at t-1 on the party vote margin at time t (using regression discontinuity) 66
2.3 Contribution limits (2013-2014) 68
2.4 Summary statistics 69
2.5 The effect of the sources of campaign contributions on the odds of a candidate winning an open-seat race between 1990 and 2014 71
2.6 Robustness: Close open-seat races 72
2.7 Robustness 74
2.8 Alternative standard errors 75
2.9 The effect of the components of campaign contributions on the odds of a candidate winning an open-seat race (based on the market structure) 76
2.10 Shorrocks-Shapley decomposition of the R-squared 77

3.1 Summary statistics* 107
3.2 The effect of ethnic heterogeneity on the number of partners* 109
3.3 The effect of ethnic hetherogeneity on the prevalence of STDs* 110
3.4 Coefficients on Ethnic Heterogeneity for the number of partners within different groups* 111
3.5 Coefficients on ethnic heterogeneity for the prevalence of STDs within different groups* 112

3.6 First-Stage Estimates* 113
3.7 IV estimates of the effect of ethnic heterogeneity on the number of partners* 114
3.8 IV estimates of the effect of ethnic heterogeneity on the prevalence of STDs* 115
3.9 Falsification: Coefficients on ethnic heterogeneity for the prevalence of non-transmitted diseases* 116

General Introduction

This thesis consists of three self-contained essays that employ a wide range of econometric tools from applied microeconomics and empirical methods to explore important issues in political and health economics. The first two chapters of this thesis investigate two main incentives underlying the campaign contributions made by different industrial sectors of the U.S. economy to political candidates. Profit seeking firms contribute to political candidates to either change the candidates' position in their favor (influence channel) or to promote their favorite candidates (electoral channel). The first chapter investigates the influence channel of contributions, while the second chapter focuses on the electoral channel. The last chapter of this thesis focuses on the effect of ethnic diversity on the prevalence of sexually transmitted diseases in Canada. The main goal of this introductory section is to provide a brief overview of the three chapters of this thesis.

Competition play a critical role in allocative efficiency and promoting long-term economic growth. A higher level of competition in the banking sector enhances innovation, choices, and efficiency which encourages development in the banking industry. A more developed financial sector moves resources to more productive activities and provides better access to the credit market, which facilitates entry and growth of the firms (Becerra et al., 2012).

In the first chapter, we investigate the extent to which campaign contributions made by banks affect the level of competition in the banking sector of the United States (influence incentive of contributions). Banks compete for deposits from customers and loans for firms. A lower level of competition in the banking sector provides banks with the opportunity to increase their profit margin by decreasing the interest rate on deposits and charging higher interest rate on loans (Allen and Gale, 2000). Therefore, banks have an incentive to influence policymakers to reduce the competition level in the financial sector. There are two main channels of influence: information andcampaign contributions. In this chapter we focus on the previously overlooked role of campaign contributions. We developed a model inspired by Grossman and Helpman (1994) and Baron (1989), to underlie the mechanism by which campaign contributions can reduce the level on competition in the banking sector.

In order to empirically test the implication of the model, we use a novel dataset which includes all contributions (not only contributions made by Political Action Committee (PACs)) made by financial institutions which are insured by Federal Deposit Insurance Corporation (FDIC) in 34 states of the United Sates between 1993 and 2010. We use Lerner index as a measure of competitiveness and estimate the effect of campaign contributions on the level of competitions using panel data analysis. Controlling for banks' characteristics, the results show contributions lead competition by one period. A one million dollar increase in lag of contributions leads to a 0.07 increase inthe Lerner index, all else equal.

Although the influence channel of contributions is not denied, similar to Snyder (1990), the focus of the second chapter is mainly on the electoral channel of contributions. Campaign contributions made by special interest groups (SIGs) has a positive impact on the vote share received by a candidate through two main channels. First, political candidates are able to use the funding raised through campaigning for advertising and informative activities and increase their probability of winning an election. Second, SIGs enforce social norms among their members by encouraging them to vote for the candidate with the most closely-aligned interests (Grossman and Helpman, 2001). In this study individuals who work for paper profit seeking firms that operate in the same market are considered special interest groups, since they have similar characteristics and share the same interest with regard to policies and regulations that affect their career.

The positive link between campaign contributions and portion of votes recieved by a candidate is well stablished. A few studies cast light on the effect of different sources of campaign contributions on candidates' probability of winning, however, this study is the first to empirically investigate the link between market share and the electoral power of firms. Firms in more concentrated industries are more likely to overcome the free-rider problem and behave as an organized interest group to advance their interests (collective action theory proposed by Olson, 1965). To empirically investigate whether concentrated firms are more successful in promoting their desire candidates, this paper uses a new detailed dataset of the elections for the U.S. House of Representatives from 1990 to 2014, which includes all individual contributions made to Political Action Committees (PACs) or candidates directly, organized by donor occupation. This paper just includes the contributions made by profit-seeking firms and categorizes the contributions based on the concentration ratio of the industry. The concentration ratio of the industry is defined as the percentage of output accounted for by the largest 50 firms of total output of the industry. The related data are obtained from the U.S. Economic Census and assigned to the contributions based on their industry codes.

A linear probability model is used to investigate the effect of market structure on the electoral power of firms. The results confirm that, even when controlling for the total value of contributions received by a candidate, political donations made by firms with high market power have a positive effect on the probability of winning. Within a race, contributions financed by more concentrated firms explain 29.54 percent of the variation in the probability of winning. This is in addition to the 25.42 percent accounted for by the total amount of contributions.

The third chapter provides evidence of a positive effect of ethnic diversity on the prevalence of sexually transmitted diseases (STDs) in Canada. Sexually Transmitted Diseases continue to be a worldwide public health concern. According to Public Health Agency of Canada (2015), rate of reported cases of chlamydia, gonorrhea, and infectious syphilis increased by 57.6, 38.9, and 101.0 percent, respectively, between 2003 and 2012. STDs not only impose a tremendous health burden on individuals, but also place a substantial economic strain on the publicly funded healthcare system of Canada.

Identifying the determinant factors that contribute to the spread of STDs is crucial for preventing future infections and efficiently allocating scare health resources. The third chapter focuses on ethnic heterogeneity as a determinant of STDs in Canada. We build on the fidelity model first proposed by Pongou (2009), to show how ethnic heterogeneity affects human sexual network formation in a multi-cultural country like Canada. Slow information flow among different ethnic groups encourages individuals to ethnically diversify their sexual partners in order to lower the chance of getting caught cheating. Anonymity thus facilitates the spread of STDs by promoting sexual infidelity and broadening sexual networks among ethnic groups.

We test the predictions of the model using a linear probability model and employing data collected from the 2013-2014 Canadian Community Health Survey (CCHS). We compute an index of ethnic heterogeneity, named HHI, at the forward sortation area level (which contains the first three characters of the postal codes) and find that ethnic diversity increases the number of sexual partners, and hence the chances of contracting a sexually transmitted disease.

Chapter 1

The Effect of Campaign Contributions on Competition in the U.S. Banking Sector

1.1 Introduction

In the 1970s, competition policy in the United States banking sector changed from exempting competition to liberalization and deregulation.[1] Competition not only plays an important role in allocational efficiency but it is also essential for long-term economic growth. Higher competition in the banking sector may stimulate the financial sector. A more developed financial sector moves resources to more productive activities and provides better access to the credit market, which facilitates entry and growth of the firms (Becerra et al., 2012). Schaeck et al. (2009), using a sample of 31 systemic banking crises in 45 countries, show that countries with more competitive banking systems have a lower risk of systemic banking crises. Although the increase in competition may erode banks' profits but banks' incentives may not necessarily be aligned with those of politicians.

[1]See Vives (2011) for a survey of competition policies and literature on banking stability and competition.

The purpose of this paper is to evaluate to what extent banks affect the level of competition in the banking sector. We focus on the banking sector in the United States. State chartered banks have incentives to reduce competition in the markets where they are operating. Among different tools to achieve this goal, we focus on the previously overlooked role of campaign contributions in decreasing competition level in the banking sector. Banks make campaign contributions to legislators who are in charge of setting policies in the banking sector.

Banks compete for both deposits from consumers and loans for firms. When competition is low, banks can provide low interest rates on deposits and charge high interest rates on loans to increase their profit margin.[2] Therefore, banks may influence policymakers to reduce competition in the banking sector. There are two main channels of influence: information and campaign contributions. In this paper we consider campaign contributions. If the marginal contributions increase banks' profits, the banks will contribute to the politician's campaign. To formalize this mechanism, we develop a model of banking contributions. The lobbying part of the model is inspired by the seminal paper by Grossman and Helpman (1994) and it shows that campaign contributions have a negative effect on the level off competition in the banking sector. In the model, the bank acts as a risk neutral matchmaker who transfers funds from consumers to entrepreneurs. There is a conflict of interest between the government and the banking sector; in the absence of lobbying, the government chooses perfect competition while the unregulated banking sector wants to monopolize the banking sector. We assume that banks are able to organize themselves in a lobbying group to make contributions to the government.

To test the model's predictions we use a new detailed dataset which includes all contributions (not only contributions made by Political Action Committee (PACs)) made by banks from 1993 to 2010 in 34 states. We use the Lerner index as a measure of competitiveness. Controlling for banks' characteristics, the results show contributions lead competition by

[2] For more information see the model presented by (Allen and Gale, 2000, Chapter 8).

one period. A one million dollar increase in lag of contributions leads to a 0.07 increase in the Lerner index, all else equal.

As a robustness check, we use different approaches to compute the Lerner index (using state dummies and computing the total cost separately for each state). The results are quite insensitive to the method of estimation of the Lerner index. In addition, we re-estimate all regressions using H-Statistics as an independent variable for a measure of competitiveness. The results confirm that contributions decrease the level of competition. Therefore, our result is robust to different methods of computing the level of competition.

The structure of the paper is as follows: the next section surveys previous literature on the impact of lobbying and campaign contributions. Section 3 introduces a model of bank lobbying. Next, we discuss data, empirical analysis and a brief overview of a competition measurement. Section 6 discusses our results and the last section presents the conclusions.

1.2 Literature review

Rajan and Zingales (2003) demonstrate that in general the existence of political connections is a crucial factor in the structure of the banking sector. They show that a new entrant reinforces competition and weakens the incumbent's opposition. As a result, incumbent banks use their political influence to set up barriers for new entrants.

Although competition is required for long-term economic growth and efficiency in the banking sector, not all countries have the same level of competition.The paper by Becerra et al. (2012) shows when there is a lot of competition in the financial market, incumbents cannot maintain the same level of rents. It points out that banks use their political connections to increase their market power, which decreases the level of competition and slows down financial development.

The Olsonian theory states that firms in more concentrated markets tend to engage more in political activities. The results from empirical studies that examine this hypothesis are mixed and no certain agreement has been made yet. (See Hansen et al. (2005) for an excellent survey.)

Specifically, there is a body lobbying literature which relates contributions to specific policies. One of the earliest study in this strand is the paper by Grossman and Helpman (1994). It introduces a model of special interests where lobbies contribute to influence the trade policies in their favour. In the model, politicians maximize the weighted sum of consumers' surplus and contributions from industries. Thus politicians provide protection to organized industries. The model has obtained empirical support (Goldberg and Maggi (1999) and Mitra et al. (2002)) and is used as a work-horse in lobbying modeling. We use this framework to describe the contributions stage in our model.

To the best of our knowledge there are only two papers which study the relationship between campaign contributions and regulation on the state level. de Figueiredo and Edwards (2007) study the effects of campaign contributions on regulatory policy decisions of state public utility commissions. Using a data set of State PAC contributions, they find that there is correlation between private money and regulatory outcomes. We go further by presenting a model which explains a causal link; contributions lead to policies. Our database is more detailed than that of de Figueiredo and Edwards (2007) since we use not only PACs' contributions but also individual contributions. As we know individual contributions comprise the majority of campaign contributions. Perez Saiz and Semenov (2015) using the same dataset as ours investigate the effect of campaign contribution on merger approval in the banking sector. They find that politicians who are in charge of bank regulation are more likely to receive contributions from banks.

More new research is studying relationships between contributions to legislators and decisions by regulatory bodies at the federal rather than state level. Mayda et al. (2010) develop a model where firms influence the government tariff decisions. Kroszner and Stratmann (1998) develop a model of interest group competition to explain how campaign contributions are donated to legislators in the financial sector.[3] They find that stronger relationships between legislators and a lower level of uncertainty increases the amount of contributions by competing groups. In our paper we do not model competition between banks directly. Instead, we consider conjectural variation approach to introduce competition level.

The implications of lobbying activity on economy, banking sector, banks and consumers have received significant interest in recent years. Igan and Mishra (2011) employ a detailed dataset which includes information about all kinds of political activities by the financial sector from 1999 to 2006. They find that lobbying activities played an important role on deregulation in the financial sector before the financial crisis in the United States. Mian et al. (2010) state that contributions had a strong effect on politician voting patterns in which politicians responded to contributions by supporting the American Housing Rescue and Foreclosure Prevention Act of 2008 and the Emergency Economic Stabilization Act of 2008.

The effect of contribution on firm-specific outcomes is unclear. There is literature which states that lobbying expenditure has a negative impact on firm performance.[4] On the other hand, there is a strand of literature that believes lobbying improves firm performance. The paper by Claessens et al. (2008) shows that Brazilian firms that made contributions to elected politicians benefited from higher stock returns and easier access to finance resources than other firms. Kim (2008), employing a dataset of S&P 500 Index firms from 1998 to 2004, shows that firms in regulated and concentrated industries are more likely to be involved in political activities, and the lobbying firm has higher equity returns relative to the market and to its industries. Chen et al. (2015), using a portfolio approach, show that there is

[3]Kroszner and Strahan (1999) use competition between interest groups to explain both state and federal bank branching deregulation in the United States.

[4]For example, an early paper by Fisman (2001) uses an event-approach to evaluate the link between politicians' connections and the downward spiral at the end of 1997 in Indonesia. The results show that firms that are politically connected are more sensitive to bad news, which led them to lose more value from 1995 to 1997 relative to those which were not connected. Bertrand et al. (2004) consider the role of CEOs' connections on corporate outcomes in France. They find that the firms which are managed by connected CEOs have relatively lower rates of return on assets and they pay lower average corporate taxes and receive higher levels of subsidies than other firms. The results also indicate that CEOs who are politically connected create more unstable jobs, and this effect is higher during the election periods.

a positive relationship between lobbying expenditure and firm performance. Faccio (2006) constructs a dataset including more than 20,000 firms from 47 countries. The results show that political connections increase the firm value and the level of political connection depends on the country characteristics.

A growing body of literature tries to investigate the lobbying influence in the banking and financial sector. Among others, Mian and Khwaja (2005) use a dataset including loan-level data of more than 90,000 firms in the financial sector of Pakistan from 1996 to 2002 to show that political connections increase the firm's loan amount and default rates. The paper by Igan et al. (2011) investigates the role of lobbying in the recent financial crisis in the United States. The results show that politically connected lenders had relatively higher levels of riskiness during 2000 to 2007, therefore they suffered more during the financial crisis. Braun and Raddatz (2009), employing a dataset including 150 countries, show that politicians who are in charge of banking regulation are more likely to be board members of banks in the future. This study, focusing on campaign contributions and the level of competition, fits better in the second body of literature.

1.3 Banks lobbying and the level of competition

To clarify the mechanism of influence of campaign contributions we provide a simple lobbying model inspired by Grossman and Helpman (1994) and Baron (1989).

1.3.1 Model

All agents is this model have access to perfect information. There are two goods in the economy: a numeraire good (consumed in the first period) and a good which is consumed in the second period. There is a continuum of consumers.

The consumers: Each consumer is endowed with ω units of numeraire goods in the first period. Consumers can consume all of the endowment in period 1 or they can save and lend it to a bank at an interest rate of r_D. Thus, consumers face the following maximization problem:

$$\max_{D} \mathbb{E}U(D) = \omega - D + U(c_2), \tag{1.1}$$

such that

$$c_2 = (1 + r_D)D \text{ (}2^{nd}\text{ period budget constraint)},$$

where D is the total amount of numeraire good that the customer deposits in period 1. $U(c_2)$ is the utility that the individual gains from consuming c_2 in the second period, where $U(c_2)$ is twice differentiable, increasing and strictly concave.

Banks: Suppose that there are n banks that compete for deposits. The degree of competitiveness of the banking sector is described by the Lerner index[5] $LI \in [0, 1]$, where $LI = 0$ corresponds to perfect competition, and $LI = 1$ is associated with a monopoly market. Banks need to satisfied certain requirements set by policymakers to be able to operate in the market, thus the degree of the competition is a choice variable for the government. We consider banks as financial intermediaries that act as matchmakers and invest the consumers deposits by lending to other customers (e.g., offering credit cards and loans). Without loss of generality, we assume that banks have access to perfect information about the type of their customers and are able to detect the borrowers that will return the loans plus the interest rate at the second period.[6] The difference between the interest rate on loans (r_L) and deposits (r_D) represents the margin of intermediation obtained by the banking sector. As in this model banks are considered as matchmakers they are not able to lend more money than what they receive as deposits from other customers. Therefor, banks face the following problem:

$$\max_{r_D, r_L, C} \Pi_b = r_L(LI)L - r_D(LI)D - C \tag{1.2}$$

[5]The Lerner is a measures of market power held by firms in a market. It captures the deviation of price from marginal cost as a fraction of price, $[\frac{(P-mc)}{P}]$, which ranges between 0 and 1.

[6]Similar to Baron (1989).

such that

$$L \leq D,$$

where $\frac{\partial r_L}{\partial LI} \geq 0$ and $\frac{\partial r_D}{\partial LI} \leq 0$. L is the total amount of loans and D represents the total amount of deposit. We assume that banks are able to organize themselves in a lobbying group and make contributions (total amount of contributions is C) to policymakers to pursue their interest.[7]

The government: The government maximizes the weighted sum of the consumers' surplus and banks' contributions C.

$$V_G = \alpha \mathbb{E} U(D) + C. \tag{1.3}$$

Note that the government does not take into account the profits of the banks. From the point of view of the government, banks' profits are the deadweight losses.[8]

Timing of the game: There are four stages, $t = 0, 1, 2, 3$.

Stage 0: The banking lobby offers contribution schedules to the government.

Stage 1: The government sets the level of competition in the banking industry. The level of competition is contingent to the contributions, $LI(C)$.[9]

Stage 2: Banks learn the level of competition and set r_D and r_L accordingly.

Stage 3: Consumers can save part of the numeraire good and lend it to the banks as deposits at the rate r_D.

Stage 4: Banks use deposits to lend to other customers.

[7] The assumption of the exogenous banking organization is reminiscent of the assumption of an organized industry in Grossman and Helpman (1994).

[8] It is possible to include the profits of banks into the government's objective function as in Grossman and Helpman (1994). The main results will be unaltered.

[9] This lobbying game is similar to the menu auction model of Bernheim and Whinston (1986).

1.3.2 The equilibrium

Consumers' first order conditions

Consider the consumers' problem as follows:

$$\max_{c_2,D}\left[\omega - D + U(c_2)\right] \tag{1.4}$$

$$\text{such that } c_2 \leq (1+r_D)D.$$

Note that because of quasi-linearity of consumers' utility and $U' > 0$ the constraint $c_2 \leq (1+y)D$ is binding. Thus to find the optimal solution of the maximization problem we substitute $c_2 = (1+r_D)D$ into the consumers' utility function and obtain a maximization problem with only one choice variable for the consumer, D. This leads to the following first-order condition:

$$1-(1+r_D)U'((1+r_D)D) = 0. \tag{1.5}$$

From this we obtain the supply function for deposits,

$$D(r_D) \equiv \frac{1}{1+r_D}\left(U'\right)^{-1}\left[\frac{1}{(1+r_D)}\right]. \tag{1.6}$$

Consumptions in the first (c_1) and second periods (c_2) are as follows:

$$c_1\left(r_D\right) = \omega - D(r_D) \text{ and} \tag{1.7}$$

$$c_2\left(r_D\right) = (1+r_D)D(r_D), \tag{1.8}$$

correspondingly.

Banks' optimum contributions solution

The banking sector takes into account the inverse demand $r_D(D)$ and $r_L(L)$ when optimizing its behaviour. Because banks only act as matchmakers, the constraint $L \leq D$ is binding. If we substitute $L = D$ in the bank's objective function we will have:

$$\max_{r_D, r_L} \Pi_b = (r_L(LI) - r_D(LI))D - C. \tag{1.9}$$

In the case that banks do not contribute anything to the policymakers the government will promotes perfect competition in the banking sector to maximize customers utility. Therefore the government facilitates the entry of new banks and promotes perfect competition in the banking sector, new banks will continue joining the banking sector to the point that the expected profit of operating in the banking sector reaches zero. Therefore, if banks operate in a perfectly competitive market (where $LI = 0$), we will have $r_L = r_D$, and hence $E(\pi_b) = 0$.

However, as the government takes into account the contributions made by banks in its objective function, banks are able to increase their intermediation margin if they contribute to the government. The first order condition of banks objective function with respect to contributions is as follows:

$$\frac{\partial \pi_b}{\partial C} = \frac{\partial r_L}{\partial LI}\frac{\partial LI}{\partial C} - \frac{\partial r_D}{\partial LI}\frac{\partial LI}{\partial C}D - 1 = 0. \tag{1.10}$$

Therefore, banks will contribute to the point that marginal gain of contributions is equal to the marginal cost of contributions.

1.3.3 Government program

At the end of stage 0 the government chooses the degree of competition in the banking industry, LI, to maximize the social welfare. We denote $c_1(LI) = c_1(r_D^*(LI))$ and $c_2(LI) = c_2(r_D^*(LI))$ the consumption levels in the equilibrium as a function of LI. The government program *without lobbying* is:

$$V_G = \max_{LI} \mathbb{E}U = c_1(LI) + U(c_2(LI)). \tag{1.11}$$

Proposition

a) The government sets $LI = 0$ *(the government wants perfect competition)*

b) The unregulated banking sector maximizes its profit at $LI = 1$*(the banking sector wants a monopoly)*

There is conflict of interest between the government and the banking sector. The government wants perfect competition in the banking sector. The banking sector wants to monopolize the market. Assume that (LI^G) is the optimal choice of the government and (LI^b) is the optimal choice of the banking sector without regulation. Then we have $LI^G = 0, LI^b = 1$.

1.3.4 The lobbying game

The lobbying game at the beginning of stage $t = 0$ is similar to Grossman and Helpman (1994).

- The bank lobby submits a contribution schedule C.
- The government selects the competition level LI.

Note that the banking lobby and the government have different objective functions. The banking lobby maximizes the utility of a representative bank

$$\pi_{b(C)}(LI(C)) - C. \tag{1.12}$$

The banking lobby has to design the contribution schedule which implements its maximum payoff given that the government obtains not less than without lobbying

$$V_G\left(LI(C)\right) \geq V(0), \tag{1.13}$$

where $LI^G = 0$ is the optimal choice of the government without lobbying. Banks contribute as long as marginal profit of contributions is greater than marginal cost of contributions:

$$\frac{\partial \pi(LI(C))}{\partial LI(C)} \frac{\partial LI(C)}{\partial C} \geq 1. \tag{1.14}$$

The government maximizes the weighted sum of consumer surplus and lobbying contributions

$$V_G(LI(C)) = \alpha \mathbb{E}U(LI(C)) + C, \tag{1.15}$$

where a schedule of competition is contingent to the contributions $LI(C)$ so the government accepts contributions as long as the marginal cost of decreasing in consumers' utility is less than the marginal gain of accepting contributions:

$$\alpha \frac{\partial \mathbb{E}U(LI(C))}{\partial LI(C)} \frac{\partial LI(C)}{\partial C} \leq 1. \tag{1.16}$$

The equilibrium will be attained when

$$\alpha \frac{\partial \mathbb{E}U(LI(C))}{\partial LI(C)} \frac{\partial LI(C)}{\partial C} = 1 = \frac{\partial \pi(LI(C))}{\partial LI(C)} \frac{\partial LI(C)}{\partial C}. \tag{1.17}$$

Equation 1.17 shows that each bank sets its level of contribution so that the marginal change in the contribution is equal to the effect of changing the level of competition on the consumers' gross welfare and to the bank's gross profit.[10]

Then we have the following(where LI^L is the level of competition in the existence of lobbying in the banking sector):

Proposition 1. $LI^G = 0 < LI^L < LI^B$

Lobbying distorts the socially efficient levels of competitiveness of the banking sector; the banking sector becomes less competitive.

[10]This result is similar to Grossman and Helpman (1994).

1.3.5 Testable implication

Our model predicts that stronger lobbying activity reduces the level of competition in the banking industry.

1.4 Data

To quantify the impact of campaign contributions on the Lerner index, this study combines annual information about banks' contributions and their characteristics in 34 states from 1993 to 2010.[11] This study includes only information about contributions that have been made by financial institutions which are insured by Federal Deposit Insurance Corporation (FDIC). More specifically, we do not include contributions that have been made by investment banks, financial associations at both regional and state level, financial consumer organizations, credit unions and thrift institutions. In this section, we discuss the data sources and the matching process in details.

State campaign contributions are obtained from data collected by the National Institute on Money in State.[12] The institute is a nonpartisan, nonprofit research organization supported by a combination of foundations and individual contributions. The institute's data include a 50-state database of contributions made from 1993 to 2014 by individuals, firms, organizations or institutions in all sectors of the United states.

This study focuses on the financial sector for the years 1993 to 2010. We consider banks as financial intermediaries who provides loans to firms using consumers deposits. Therefore, financial institutions are excluded from the data if they do not have both consumer loans and business loans at the same time, as they have a different Business model (334 banks

[11] At the time of collecting and cleaning the dataset, data were available up until 2010. We identified contributions of 34 states. We are planning to cover the rest in the future studies. The following states are included in our dataset: Alaska, Alabama,Arkansas, Arizona, California, Colorado, Connecticut, District of Columbia,Delaware, Florida, Hawaii, Iowa, Idaho, Indiana, Kansas, Kentucky, Maine, Minnesota, Mississippi, Montana, North Dakota, Nebraska, New Jersey, New Mexico, Nevada, Oregon, Rhode Island, South Dakota, Texas, Utah, Virginia, Vermont, West Virginia, Wyoming.

[12] Data are available at www.followthemoney.org.

are excluded). We also exclude if the institutions have reported a negative income (In total, 544 banks, have reported a negative income). The first or/and the last year of operation are excluded form the dataset, in the case that the bank has been started operation or has been stopped working after 1993 (In total 1249 records are excluded.). In this case there is a dramatic change in the income and the cost of the bank in the first or last year of operation and other years of operation.

We not only consider contributions made by banks, including contributions made by banks' executives. We use the bank identifiers from the Call Reports in order to match the individual contributions in the financial sector to the unique institution FDIC certificate code. We use geographic location such as the state name, the city name, the postal code and address, the name of contributor, the name of affiliated companies and the year of contribution to identify the contributor in the Call Reports. In some circumstances, to ensure of the accuracy of the matching process, it was necessary to use Internet sources such as city-data, campaignmoney, LinkedIn and bank websites. In the end, the study was able to match more than 80% of contributions successfully.[13] Table 1.1 presents the mean of contributions made by bank from 1993 to 2010. The mean of contributions is significantly smaller in non-election years. This is also true for the number of contributors.

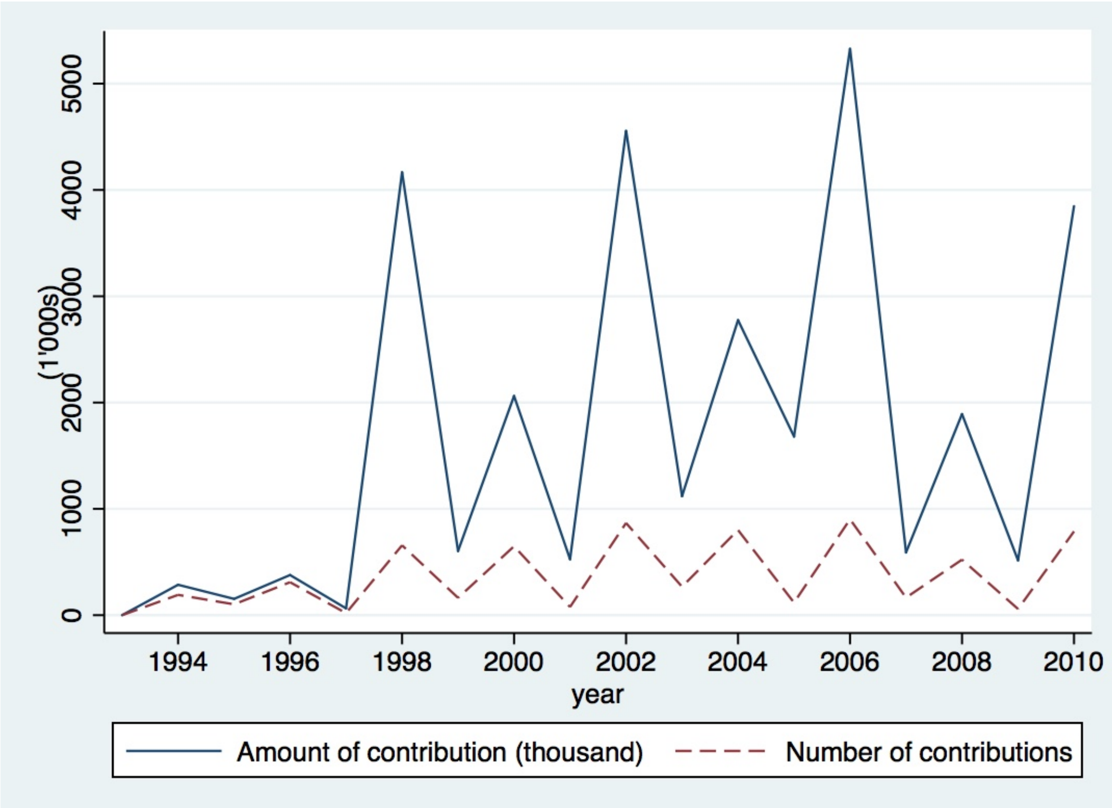
Figure 1.1 presents the trend in campaign contributions in the financial sector from 1993 to 2010. The solid line represents the total amount of donations (in thousand of dollars) made by financial institutions in each year, and the dashed line shows the number of contributors in each year. Figure 1.1 shows that financial institutions tend to contribute more in election years (Most of the states have elections in even years).

In order to obtain information about individual bank characteristics, this study uses the Consolidated Report of Condition and Income (referred as the Call Report).[14] All regulated

[13]There was not enough information available to match all the contributions. More specifically, some contributors did not provide any information about the name of contributor, postal code, address and the name of affiliated company. Therefore, it was not possible to identify and link them to a financial institutions. In total, 114,301 records of contributions are available from 1993 to 2010 in 34 states, and we were able to identify 94,830 records successfully.

[14]Data are available at https://www2.fdic.gov/sdi/download_large_list_outside.asp.

Figure 1.1: The trend of contributions by bank from 1993 to 2010



Contributions are in thousand of dollar.
The peaks are election years in the United States.

financial institutions in the U.S. are required by the Federal Financial Institutions Examination Council (FFIEC) to report their financial and all other basic information in the unique format of Call Report quarterly, and the Federal Deposit Insurance Corporation (FDIC) is responsible for collecting, overseeing and storing Call Reports of all insured financial institutions.[15] For the purpose of this study, we use annual Call Reports to be able to match them with contributions data which were available only annually. Each bank has a unique FDIC certificate code which permits the identification of a bank over years of operation, because in some cases the name of the bank has been changed more than once. Each Call Report has 63 different categories, and for the purpose of this study we use the following categories: Performance and Condition Ratios, income and expenses, net loans and leases, assets and

[15]https://www.fdic.gov/regulations/resources/call/index.html.

liabilities and demographics information.

Table 1.1 presents summary statistics for the main variables used in this paper. As shown in Panel B, there are 96,224 observations in our dataset covering 8,837 financial institutions between 1993 to 2010 in the United States. On average 328 financial institutions make a donation on election years. A financial institution on average contributes 559 (82.68) dollars to political candidates in an election year (a non-election year). The mean of Lerner index is about 0.6, which means the banking industry operates under monopolistic competition (the average value of H-statistics, which is equal to 0.7, supports this claim).

1.5 Empirical analysis

The above discussed model implies that banks in a form of organized interest group use their political influence to monopolize the market in order to extract more rents in the market. Therefore the model predicts that contributions decrease the competition level. The following section presents an econometric model to investigate the link between contributions and competition in the financial sector. We use Lerner index as a measure of competitiveness.

1.5.1 The Lerner index as a measure of competitiveness

Among different measures of competitiveness,[16] HHI, H-Statistics and Lerner Index are more popular in the literature of studying banking industry. Bolt and Humphrey (2012) show that these measures are uncorrelated and measure-specific for the U.S. banking industry. H-Statistic measures changes in total revenues resulting from changes in input prices, $H = \sum_{k=1}^{m} \frac{\partial R}{\partial P_k} \frac{P_k}{R}$, which lies between minus infinity to unity; a negative H-statistic shows the market is a monopoly or a perfect colluding oligopoly. While the Lerner index is of one the most universally used measures of market power held by firms in a market,[17] and it is also called the price-cost margin. It captures the deviation of price from marginal cost as a fraction of

[16]See Carlton and Perloff (1990) Chapter 8 for an excellent survey of the measures of competitiveness.
[17]It was developed by Lerner in 1934.

price $[\frac{(P-mc)}{P}]$ and ranges between 0 and 1. When price is equal to marginal cost, the index is zero which means the firm has no market power to increase price beyond marginal cost. The larger the index, the lower level of competition. If there is only one firm operating in a market, the index is one, which indicates a monopoly situation.

For a profit maximizing firm the Lerner index is equal to the inverse elasticity of demand curve that firm faces in the market $(\frac{1}{\epsilon})$. If the Lerner index is greater that zero, it indicates that the price is higher than the marginal cost and it is a sign of positive economic profits in that market. Therefore, if there are no barriers in the market new firms are willing to enter that market in order to earn profit. Presence of the new entrant increases the elasticity of the incumbents' demand curve and reduces their market power, and as a result, the Lerner index.

The choice of proper measure of competition is highly controversial; there is no best or most accurate measure of market power.[18] Bikker and Haaf (2002) and Carlton and Perloff (1990) state that most of the academic literature in the banking sector has a strong preference for using the Lerner index as a measure of competitiveness.[19] We use the Lerner index as a measure of competitiveness for three reasons. First, computing the Lerner index does not require defining the relevant market, which allows estimation for each bank separately and it provides the possibility of estimating the effect of contributions on the level on competition. Second, the Lerner index carries information about the evolution of the elasticity of the demand that the financial institution faces which provides information about the structure of the market which the firm is operating. Third, Tirole (1988) shows that under certain assumptions the Lerner index is related to measures of welfare losses, therefore, we will be able to compute the welfare effect of contributions.

The first step of computing the Lerner index is estimating the marginal cost of each bank. We use a translog function to estimate the cost function of each bank. Benston et al.

[18]For more information see Martin (1993).

[19]For instance, look at are Shaffer (1993) for Canadian banks, Angelini and Cetorelli (2003) for Italian banks, Fernandez de Guevara et al. (2005) for the European banking sector, Maudos and Nagore (2005), Jiménez et al. (2013).

(1982) were the first to use the translog form to estimate total cost in the banking sector. Many other papers have since adopted this approach.[20] Although all the papers follow the same general format, they differ, however, in the definition of outputs and inputs in the banking sector. The definition that we use in this paper is close to the one used by Bolt and Humphrey (2012):

$$\begin{aligned}\ln TC_{bt} = \\ &\alpha_0 + \sum_{i=1}^{3} \alpha_i \ln Q_{ibt} + \sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_{ij} (\ln Q_{ibt} \ln Q_{jbt}) \\ &+ \sum_{i=1}^{3}\sum_{k=1}^{3} \delta_{ij} (\ln Q_{ibt} \ln P_{kbt}) + \sum_{k=1}^{3} \beta_k \ln P_{kbt} + \sum_{k=1}^{3}\sum_{m=1}^{3} \beta_{km} (\ln P_{kbt} \ln P_{mbt}) \\ &+ \theta_1 \ln LLR_{bt} + \theta_2 (\ln LLR_{bt})^2 + \theta_3 \ln T + \theta_4 (lnT)^2,\end{aligned} \tag{1.18}$$

where TC_{bt} is the total operating and interest expenses of financial institution b at time t, Q_{ibt} is the value of three outputs: the value of the level of consumer loans, the value of the level of business loans and the value of the level of securities held by the financial institution b at time t. P_{kpt} is the value of three input prices: P_{Ft}, the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, the annual average expense per full-time equivalent worker which is defined as the ratio of salaries and benefits to the number of full-time equivalent employees; and P_{Kt}, an approximation to the cost of capital which is defined as the ratio of premises expense to the value of premises. LLR_{bt} is the loan loss ratio of the financial institution at time t which is calculated as (value of all loans and leases - provision for losses on loan and leases)/(value of all loans and leases). T is a quarterly time indicator variable used to reflect time-related technical change.

[20]For a few examples among many see Angelini and Cetorelli (2003); Koetter et al. (2012); Berger et al. (2009); Fernandez de Guevara et al. (2005) and; Berg and Kim (1994).

The translog form is estimated using a panel model. Marginal cost is

$$MC_{bt} = (\sum_{i=1}^{3} d\ln TC_{bt}/d\ln Q_{ibt})AC_{bt}, \tag{1.19}$$

where AC reflects total average cost which is equal to the ratio of total cost to total asset $AC = TC/TA$. The Lerner index is defined as $(P - MC)/P$ where price is equal to the ratio of total revenue to total asset $P = TR/TA$. Table 1.1 presents the mean and the standard deviation of the estimated Lerner index.

1.5.2 The effect of campaign contributions on the Lerner index

To estimate the effect of campaign contributions on the Lerner index, we use a sample of 34 states from 1993 to 2010 to estimate the following fixed effects model:

$$\begin{aligned} Lerner_{bt} = & \\ & \alpha_0 + \alpha_1 Cont_{bt} + \alpha_2 Cont_{bt-1} + \alpha_3 Election_{bt} \\ & + \alpha_4 Asset_{bt} + \alpha_5 Roa_{bt} + \alpha_6 NumEmp_{bt} \\ & + \alpha_7 OffDom_{bt} + \alpha_8 MulStat_{bt} + \alpha_9 EffDate_{bt} \\ & + \sum_{c=1} \theta_c BnkCls_{cbt} + \sum_{a=1} \gamma_a AssCon_{abt} + \sum_{i=1} \lambda_i InsMem_{ibt} \\ & + \delta_s + \psi_t + \epsilon_{bt}, \end{aligned} \tag{1.20}$$

where $Lerner_{bt}$ is the value of the Lerner index for bank b at time t; $Cont_{bt}$ and $Cont_{bt-1}$ are the sums of money contributed by the institution to politicians in year t and $t-1$, respectively; $Election_{bt}$ is a dummy which is one in the election years (and zero otherwise), which captures the trends of contributions in election years shown in Figure 1.1.

We also control for time variant bank characteristics using the following variables: $Asset_{bt}$ is the sum of all assets owned by the institution; Roa_{bt} is the return of assets; $NumEmp_{bt}$ is the number of full-time employees of the institution; $OffDom_{bt}$ is the number of domestic offices operated by the institution; $MulStat_{bt}$ is a dummy which is one when the institution can accept deposits in more than one state; $EffDate_{bt}$ is a dummy whose value is one for the year in which a structural change was implemented by the institution; $BnkCls_{cbt}$ is a vector of dummies which includes dummies for the institution's charter class which is assigned by FDIC;[21] $AssCon_{abt}$ is a vector of dummies which contains dummies of the institution's primary specialization regarding of asset concentration at time t;[22] $InsMem_{ibt}$ is a subset of dummies including the institution's insurance fund membership type at time t.[23]

δ_s, and ψ_t are state, and time fixed effects respectively. δ_s are controls for time invariant unobservable factors within a states that have an impact on the Lerner index of banks in that state. ψ_t includes dummies for years to capture trends in contributions and the competition level. Finally, ϵ_{bt} is the error term.

1.6 Results

Table 1.2 reports the results of estimation Equation (1.18). Column 1 shows the baseline Panel results with fixed effects and Column 2 shows Panel results of estimation of the same equation using fixed effects and state dummies. The rest of Table 1.2 illustrates the results of estimation Equation (1.18) using bank and state dummies separately for each state. The results of the estimations are used for computing the Lerner index based on Equation (1.19).[24]

[insert Table 1.2 here]

[21]This code is assigned based on the institution's charter type, charter agent, Federal Reserve membership status and primary federal regulator.

[22]Based on FDIC, there are eight different types of asset concentration: international specialization, agricultural specialization, credit-card specialization, commercial lending specialization, mortgage lending specialization, consumer lending specialization, other specialized < \$1 billion, all other < \$1 billion, and all other > \$1 billion.

[23]Based on FDIC, there are three different types of membership: Deposit Insurance Fund, Bank Insurance Fund, Savings Association Insurance Fund.

[24]In order to examine the significant difference between the states, the equation is estimated separately for each state. The results are available in Table 1.2.

The main results are summarize in Tables 1.3 and 1.4, presenting the effect of campaign contributions on the competition level in the banking industry using different sets of controls. Table 1.3 reports the estimated effect of results of campaign contributions on the Lerner index based on Equation (1.20). All the error terms in this study are clustered at the institution level to account for serial correlation within the group. Column 1 shows that after controlling for all time invariant bank-specific characteristics (by using bank fixed effects), a one million dollars increase in lag of contributions leads to a 0.069 increase in the Lerner index, all else equal. This effect is very large considering that the Lerner index lies between zero and one and the average value of the Lerner index is equal to 0.594. While the sum of lagged contributions is positively and significantly correlated with the Lerner index, the sum of contributions itself is not significant. This result suggests that it takes time to see the effect of contributions or in the other words, the contributions lead the Lerner index by one period. The point estimate of the lag of contributions drops to 0.054 when we add state dummies (Column 2), and the estimated coefficient of the contributions declines and stays insignificant. We recalculate the Lerner index separately for each state and rerun the estimation with the new Lerner index. The results of this experience is shown in Column 3. As shown the point estimate of the lag of contributions increases slightly to 0.056, which suggests that the results of the paper is not driven by a certain specification used to compute the Lerner index.

[insert Table 1.3 here]

1.7 Robustness checks

Table 1.4 shows the results of the same model where a second lag of contributions is added. The point estimate of the second lag of contributions is positive but statistically insignificant. More importantly, adding the second lag of contributions does not affect the significance of the point estimates and the magnitudes change slightly. Taking into account that the United

States has state level elections every two years (which are the even years in most states) and as discussed before most of the contributions are made in the year of elections, this result suggests that the effect of contributions for a given political cycle will manifest themselves before the next election cycle and after the next election the effect does not longer exist. This result is consistent with the finding of the paper by Chen et al. (forthcoming).

To further gauge the robustness of the regression results, this paper uses H-Statistics as an alternative measure of competitiveness. H-Statistic is developed by Panzar and Rosse (1987), representing changes in total revenues resulting from changes in input prices $H = \sum_{k=1}^{m} \frac{\partial R}{\partial P_k} \frac{P_k}{R}$.[25] H-Statistic lies between minus infinity to unity; a negative H-statistic shows the market is a monopoly or a perfect colluding oligopoly. Under the perfect competition market, H-Statistic equals to unity and if the market structure is characterized by monopolistic competition, the H-Statistic ranges between zero and one.

Table 1.6 illustrates the results of the estimation of the total revenue based on Equation (1.21). The first column shows the results of the baseline specification, in which only time fixed effects are used. Column 2 illustrates the results of the same estimation when state dummies are added. The Equation (1.22) is used for calculating H-Statistics, and re-run the estimations to see the effect of contributions on H-Statistic as an alternative measure of concentration. The results of this practice are summarized in Table 1.5. Column 1 illustrates that a one million dollars increase in lag of contributions leads to a 0.049 unites decrease in the H-Statistic when only time and bank fixed effects are used. Column 2 shows the point estimates of the lag of contributions increases slightly to -0.0489 when state dummies are added. However, the magnitude of Lerner index and H-statistics are not comparable, as they measure the concentration of the market in different ways. However, the results are suggestive that, regardless of the measure of concentration of the market, more contributions will lead to less competition in the next period. These results are consistent with the prediction of the model and the main empirical results.

[25] H-Statistic calculation is shown in the appendix.

[insert Table 1.5 here]

1.8 Conclusion

Competition plays an important role in the banking sector; it is not only the key factor of allocation efficiency but also a crucial factor in long-term economic growth. This paper develops a model to show the influence of banks' contributions on the competition level. In the absence of lobbying activities, the government chooses perfect competition while an unregulated bank wants to monopolize the market in order to extract the maximum level of rent. The model predicts that contributions are truthful and contributions reduce competition in the financial sector.

This paper employs a novel detailed data set which includes campaign contribution by each bank to investigate the effect of lobbying on the Lerner index (as a measure of competition level in the banking industry). The results show that, after controlling for bank level variables, the effect of contributions is not instant and in fact contributions lead competition by one period. Results show that a higher level of contributions increases the Lerner index or in there other words, decreases the level of competition. We also find that the second lag of contributions does not have a significant effect on the Lerner index. It suggest that the effect of contributions for a given political cycle will manifest themselves before the next election cycle and after the next election the effect does not longer exist. This paper uses H-Statistics as another measure of competition to gauge the robustness of the results. The results are consistent with the prediction of the model.

Appendix

Estimation of H-Statistic

There are several papers that apply Panzar and Rosse (1987) model in the banking sector.[26] We use a translog function in order to calculate H-Statistic, the method we apply in this paper is close to the paper by Bolt and Humphrey (2012).

$$\begin{aligned}
\ln TR_{bt} = \\
& \alpha_0 + \sum_{i=1}^{5} \alpha_i \ln Q_{ibt} + \sum_{i=1}^{5}\sum_{j=1}^{5} \alpha_{ij} (\ln Q_{ibt} \ln Q_{jbt}) \\
& + \sum_{i=1}^{5}\sum_{k=1}^{3} \delta_{ij} (\ln Q_{ibt} \ln P_{kbt}) + \sum_{k=1}^{3} \beta_k \ln P_{kbt} + \sum_{k=1}^{3}\sum_{m=1}^{3} \beta_{km} (\ln P_{kbt} \ln P_{mbt}) \\
& + \theta_1 T + \theta_2 T^2 + \sum_{i=1}^{5} \rho_i T \ln Q_{ibt} + \sum_{k=1}^{3} \lambda_k T \ln P_{kbt},
\end{aligned} \tag{1.21}$$

where TR_{bt} is the value of total revenue for financial institution b at time t; Q_{bt} is the value of three outputs for financial institution b at time t:(the value of the level of consumer loans, the value of the level of Business loans and the value of the level of securities held by bank)plus LLR (which is defined is the text)and the ratio of total domestic plus foreign deposits to total liabilities; P_{kbt} is the value of the same three output prices which are defined in the text; T is the same time indicator.

The H-Statistic is driven from the following equation:

$$H_{bt} = \sum_{k=1}^{3} d \ln TR_{bt} / d \ln P_{kbt}, \tag{1.22}$$

[26] For a few examples among many see Molyneux et al. (1994); Bikker and Haaf (2002); Claessens and Laeven (2004); Schaeck and Cihak (2012).

Tables

Table 1.1: Summary statistics from 1993 to 2010

Variables	Mean	Standard Dev.
Panel A		
Contributions in Election Years	559.25	8,376.62
Number of Contributioner in Election Years(per year)	327.77	
Contributions in Non-Election Years	82.68	5,452.61
Number of Contributioner in Non-Election Years(per year)	42.94	
Panel B		
Lerner Index	0.594	0.0849
H-statistics	0.708	0.0831
Asset(thousands of dollars)	553,121.6	7,881,879
ROA	0.984	1.063
Consumer Loans(thousands of dollars)	71,566.42	1,421,348
Business Loans(thousands of dollars)	55,440.04	888,554.6
Total Deposits(thousands of dollars)	382,025.8	5,175,016
Total Revenue(thousands of dollars)	44,174.17	552,262.3
Total Cost(thousands of dollars)	31,073.18	378,461.5
Number of Employees	139.668	1,421.141
Number of Domestic Offices	7.158	37.491
Number of Observations	96,224	
Number of banks	8,837	

Table 1.2: Estimation of total cost of banks

	ln(TC) (1)	ln(TC) (2)	ln(TC_{AK}) (3)	ln(TC_{AL}) (4)	ln(TC_{AR}) (5)	ln(TC_{AZ}) (6)	ln(TC_{CA}) (7)	ln(TC_{CO}) (8)	ln(TC_{LA}) (9)
$ln(ConsumerLoans)$	0.0819*** (0.0094)	0.0798*** (0.0094)	2.0784*** (0.4505)	0.1089 (0.0934)	0.6936*** (0.0900)	0.7015*** (0.1875)	0.0599* (0.0345)	0.2741*** (0.0737)	-0.0815 (0.0671)
$ln(BusinessLoans)$	-0.2358*** (0.0067)	-0.2371*** (0.0067)	-1.5656*** (0.2812)	-0.5247*** (0.0593)	-0.6128*** (0.0718)	0.0459 (0.2133)	-0.2268*** (0.0290)	-0.6627*** (0.0570)	-0.2890*** (0.0296)
ln(Securities)	0.1633*** (0.0083)	0.1625*** (0.0083)	1.7983*** (0.2769)	0.6989*** (0.0737)	-0.0479 (0.0554)	0.0755 (0.1087)	0.0993*** (0.0282)	0.2929*** (0.0504)	0.4375*** (0.0541)
$ln(ConsumerLoans)\times ln(ConsumerLoans)$	0.0184*** (0.0003)	0.0185*** (0.0003)	-0.0123 (0.0221)	0.0476*** (0.0059)	0.0128*** (0.0040)	0.0183*** (0.0058)	0.0043*** (0.0008)	0.0031 (0.0034)	0.0220*** (0.0030)
$ln(BusinessLoans)\times ln(BusinessLoans)$	0.0205*** (0.0002)	0.0205*** (0.0002)	-0.0011 (0.0038)	0.0194*** (0.0012)	0.0164*** (0.0013)	0.0436*** (0.0036)	0.0158*** (0.0007)	0.0352*** (0.0014)	0.0103*** (0.0007)
$ln(Securities)\times$ln(Securities)	0.0191*** (0.0002)	0.0191*** (0.0002)	-0.0025 (0.0084)	0.0565*** (0.0030)	0.0221*** (0.0012)	0.0132*** (0.0019)	0.0128*** (0.0006)	0.0121*** (0.0009)	0.0163*** (0.0009)
$ln(ConsumerLoans)\times ln(BusinessLoans)$	-0.0097*** (0.0003)	-0.0096*** (0.0003)	0.0494** (0.0191)	0.0056 (0.0053)	0.0100** (0.0046)	-0.0519*** (0.0070)	-0.0035*** (0.0012)	0.0069** (0.0030)	-0.0021 (0.0019)
$ln(ConsumerLoans)\times$ln(Securities)	-0.0087*** (0.0004)	-0.0087*** (0.0004)	-0.0610*** (0.0146)	-0.0841*** (0.0074)	0.0093** (0.0046)	0.0042 (0.0053)	0.0023** (0.0012)	0.0016 (0.0033)	-0.0186*** (0.0043)
$ln(BusinessLoans)\times$ln(Securities)	-0.0074*** (0.0003)	-0.0074*** (0.0003)	-0.0086 (0.0170)	-0.0019 (0.0046)	-0.0154*** (0.0037)	-0.0154*** (0.0054)	-0.0035*** (0.0011)	-0.0083*** (0.0026)	0.0004 (0.0016)
$ln(P_l)$	0.0722* (0.0389)	0.0676* (0.0390)	8.5755*** (2.4044)	-0.4006 (0.2705)	0.2858 (0.2936)	0.5809 (0.8208)	0.8438*** (0.2424)	-0.1268 (0.2764)	0.0873 (0.2187)
$ln(P_f)$	1.1158*** (0.0277)	1.1153*** (0.0277)	1.2784* (0.6656)	0.0231 (0.1571)	1.4604*** (0.1847)	1.1887** (0.5752)	1.4226*** (0.1245)	1.1953*** (0.1555)	1.3270*** (0.1453)
$ln(P_k)$	0.2851*** (0.0153)	0.2846*** (0.0153)	-0.5484 (0.6097)	1.2333*** (0.1063)	0.5255*** (0.1050)	-0.3144 (0.3115)	-0.0544 (0.0777)	0.3726*** (0.0935)	1.0320*** (0.0928)
$ln(ConsumerLoans)\times(P_l)$	-0.0114*** (0.0019)	-0.0110*** (0.0019)	-0.3416*** (0.0825)	-0.0358* (0.0210)	-0.2073*** (0.0176)	-0.0870*** (0.0331)	-0.0012 (0.0076)	-0.0439*** (0.0120)	0.0207 (0.0134)
$ln(BusinessLoans)\times(P_l)$	0.0410*** (0.0014)	0.0411*** (0.0014)	0.3081*** (0.0538)	0.0643*** (0.0130)	0.0889*** (0.0131)	0.0450 (0.0361)	0.0215*** (0.0062)	0.0549*** (0.0110)	0.0429*** (0.0052)
$ln(Securities)\times(P_l)$	-0.0307*** (0.0018)	-0.0308*** (0.0018)	-0.2348*** (0.0540)	-0.1093*** (0.0156)	0.0061 (0.0120)	-0.0172 (0.0176)	-0.0300*** (0.0063)	-0.0791*** (0.0096)	-0.0898*** (0.0102)
$ln(ConsumerLoans)\times(P_f)$	0.0062*** (0.0013)	0.0065*** (0.0013)	0.0369 (0.0340)	-0.0281* (0.0150)	0.0385*** (0.0117)	0.0065 (0.0229)	0.0154*** (0.0042)	0.0362*** (0.0096)	-0.0078 (0.0093)

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	ln(TC) (1)	ln(TC) (2)	$\ln(TC_{AK})$ (3)	$\ln(TC_{AL})$ (4)	$\ln(TC_{AR})$ (5)	$\ln(TC_{AZ})$ (6)	$\ln(TC_{CA})$ (7)	$\ln(TC_{CO})$ (8)	$\ln(TC_{LA})$ (9)
$ln(BusinessLoans)\times(P_f)$	-0.0193***	-0.0195***	-0.0332	-0.0350***	-0.0797***	0.0284	-0.0220***	-0.0419***	-0.0191***
	(0.0010)	(0.0010)	(0.0249)	(0.0095)	(0.0102)	(0.0218)	(0.0048)	(0.0079)	(0.0047)
$ln(Securities)\times(P_f)$	0.0210***	0.0207***	-0.0078	0.0962***	0.0266***	0.0038	0.0176***	-0.0008	0.0304***
	(0.0011)	(0.0011)	(0.0207)	(0.0116)	(0.0085)	(0.0123)	(0.0037)	(0.0067)	(0.0062)
$ln(ConsumerLoans)\times(P_k)$	0.0014*	0.0014*	0.0520	-0.0650***	-0.0017	0.0049	-0.0050	-0.0141**	-0.0032
	(0.0008)	(0.0008)	(0.0387)	(0.0086)	(0.0096)	(0.0151)	(0.0031)	(0.0070)	(0.0063)
$ln(BusinessLoans)\times(P_k)$	-0.0104***	-0.0103***	0.0016	0.0108**	-0.0101	0.0198	-0.0036	-0.0034	-0.0064**
	(0.0006)	(0.0006)	(0.0284)	(0.0055)	(0.0071)	(0.0142)	(0.0027)	(0.0057)	(0.0029)
$ln(Securities)\times(P_k)$	-0.0017**	-0.0019**	-0.0030	0.0260***	-0.0053	-0.0050	-0.0070***	0.0004	-0.0096*
	(0.0007)	(0.0007)	(0.0312)	(0.0076)	(0.0052)	(0.0072)	(0.0027)	(0.0043)	(0.0053)
$ln(P_l)\times(P_l)$	-0.0191***	-0.0188***	-0.8479***	0.1577***	0.0579*	-0.0103	-0.0887***	0.1051***	0.0202
	(0.0037)	(0.0037)	(0.2532)	(0.0312)	(0.0314)	(0.0792)	(0.0260)	(0.0239)	(0.0232)
$ln(P_f)\times(P_f)$	0.0466***	0.0463***	-0.0109	-0.0256***	0.0387**	0.1373***	0.0975***	0.0408***	0.0989***
	(0.0023)	(0.0024)	(0.0343)	(0.0084)	(0.0181)	(0.0499)	(0.0118)	(0.0119)	(0.0131)
$ln(P_l)\times(P_k)$	-0.0502***	-0.0500***	-0.0107	-0.2458***	-0.0893***	-0.0203	0.0306*	-0.0659***	-0.2243***
	(0.0033)	(0.0033)	(0.1135)	(0.0212)	(0.0247)	(0.0539)	(0.0176)	(0.0181)	(0.0197)
$ln(P_k)\times ln(P_k)$	0.0040***	0.0039***	-0.0037	0.0224***	-0.0259***	0.0058	-0.0428***	-0.0151**	0.0126**
	(0.0008)	(0.0008)	(0.0211)	(0.0066)	(0.0057)	(0.0168)	(0.0039)	(0.0059)	(0.0055)
$ln(P_l)\times(P_f)$	-0.1143***	-0.1141***	-0.2668*	-0.0403	-0.1511***	-0.0579	-0.0923***	-0.1027***	-0.0609**
	(0.0049)	(0.0049)	(0.1597)	(0.0296)	(0.0330)	(0.0796)	(0.0227)	(0.0298)	(0.0247)
$ln(P_f)\times(P_k)$	0.0067***	0.0070***	-0.0555*	0.0390***	0.0499***	-0.0643**	-0.0071	0.0185	0.0200*
	(0.0023)	(0.0023)	(0.0324)	(0.0143)	(0.0155)	(0.0324)	(0.0103)	(0.0118)	(0.0117)
$ln(LoanLossRation)$	-1.2855***	-1.2820***	11.7811***	-2.1497***	-0.6886*	-0.6156	-1.5420***	-2.6908***	-1.9715***
	(0.0618)	(0.0617)	(3.9035)	(0.5690)	(0.3650)	(1.1739)	(0.2405)	(0.6229)	(0.5384)
$ln(LoanLossRation)^2$	-0.2664***	-0.2654***	396.3881**	0.8740	-0.0583	-1.0266	2.8440***	-5.8738	6.5041
	(0.0213)	(0.0213)	(170.4748)	(8.0217)	(0.9254)	(2.3792)	(0.6197)	(3.8171)	(9.0894)
$(TimeIndicator)$	-0.1567***	-0.1567***	-0.0812	-0.0916***	-0.2004***	-0.3395***	-0.3513***	-0.1690***	-0.1584***
	(0.0037)	(0.0037)	(0.0638)	(0.0163)	(0.0174)	(0.0757)	(0.0186)	(0.0210)	(0.0154)
$(TimeIndicator)^2$	0.1032***	0.1032***	0.1042***	0.0681***	0.1117***	0.2631***	0.2040***	0.1239***	0.1050***
	(0.0014)	(0.0014)	(0.0258)	(0.0062)	(0.0066)	(0.0275)	(0.0074)	(0.0081)	(0.0057)
State dummies	N	Y	N	N	N	N	N	N	N
Number of Observations	96,224	96,224	139	3,037	3,388	661	5,475	3,223	3,251
Number of Banks	8,837	8,837	10	274	294	95	659	373	281
R-Squared	0.775	0.776	0.936	0.874	0.871	0.845	0.774	0.851	0.825

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{CT})$ (10)	$\ln(TC_{DC})$ (11)	$\ln(TC_{DE})$ (12)	$\ln(TC_{FL})$ (13)	$\ln(TC_{HI})$ (14)	$\ln(TC_{IA})$ (15)	$\ln(TC_{IN})$ (16)	$\ln(TC_{KS})$ (17)	$\ln(TC_{KY})$ (18)
$ln(ConsumerLoans)$	0.6113***	1.1926***	0.6522***	0.3363***	-0.0755	0.3470***	-0.0921	0.1115**	0.3599***
	(0.1094)	(0.4372)	(0.0968)	(0.0485)	(0.2971)	(0.0513)	(0.0675)	(0.0519)	(0.0638)
$ln(BusinessLoans)$	-0.3438***	-0.7909*	0.2011**	-0.2067***	-0.0618	-0.6666***	-0.2685***	-0.3429***	-0.4722***
	(0.0504)	(0.4005)	(0.0914)	(0.0346)	(0.1020)	(0.0404)	(0.0346)	(0.0338)	(0.0297)
ln(Securities)	0.1683*	-1.2007*	-0.2474**	0.1420***	0.2358	0.3083***	0.0833*	0.3940***	-0.0551
	(0.0908)	(0.6345)	(0.1246)	(0.0389)	(0.3099)	(0.0312)	(0.0462)	(0.0374)	(0.0541)
$ln(ConsumerLoans)\times ln(ConsumerLoans)$	-0.0038	0.0167*	0.0210***	0.0125***	0.0517***	0.0037	0.0192***	0.0056**	0.0200***
	(0.0033)	(0.0092)	(0.0022)	(0.0014)	(0.0154)	(0.0025)	(0.0023)	(0.0024)	(0.0020)
$ln(BusinessLoans)\times ln(BusinessLoans)$	0.0102***	0.0279***	0.0096***	0.0164***	0.0022	0.0222***	0.0129***	0.0214***	0.0153***
	(0.0010)	(0.0063)	(0.0020)	(0.0008)	(0.0020)	(0.0008)	(0.0007)	(0.0007)	(0.0008)
$ln(Securities)\times$ln(Securities)	0.0117***	0.0705***	0.0186***	0.0243***	0.0177**	0.0196***	0.0165***	0.0221***	0.0195***
	(0.0022)	(0.0241)	(0.0022)	(0.0009)	(0.0089)	(0.0006)	(0.0008)	(0.0007)	(0.0009)
$ln(ConsumerLoans)\times ln(BusinessLoans)$	0.0036	0.0107	-0.0134***	-0.0038**	-0.0053	0.0077***	-0.0052***	0.0086***	-0.0005
	(0.0025)	(0.0197)	(0.0023)	(0.0016)	(0.0074)	(0.0024)	(0.0016)	(0.0024)	(0.0020)
$ln(ConsumerLoans)\times$ln(Securities)	0.0028	-0.0927***	-0.0184***	-0.0236***	-0.0513**	-0.0129***	-0.0039	-0.0025	-0.0186***
	(0.0053)	(0.0259)	(0.0038)	(0.0022)	(0.0235)	(0.0023)	(0.0027)	(0.0031)	(0.0037)
$ln(BusinessLoans)\times$ln(Securities)	-0.0085***	-0.0016	-0.0038	-0.0007	0.0137*	-0.0055***	-0.0053***	-0.0127***	0.0073***
	(0.0029)	(0.0291)	(0.0026)	(0.0016)	(0.0072)	(0.0014)	(0.0014)	(0.0018)	(0.0016)
$ln(P_l)$	-0.8373	-4.6545**	0.5503	-0.9697***	-0.2102	0.4116***	-0.5142*	0.6225***	0.5268**
	(0.5573)	(2.1309)	(0.7599)	(0.2121)	(1.1092)	(0.1579)	(0.2705)	(0.1634)	(0.2095)
$ln(P_f)$	1.3265***	-1.6717	-1.2169**	1.3463***	1.1922***	1.0410***	2.3070***	1.0555***	0.9854***
	(0.3230)	(1.2923)	(0.4771)	(0.1381)	(0.3761)	(0.1042)	(0.1968)	(0.1142)	(0.1512)
$ln(P_k)$	0.2160	0.2780	-0.1966	0.4671***	-0.1556	0.0898*	-0.1578	0.1374**	0.1707**
	(0.1849)	(0.7682)	(0.2973)	(0.0727)	(0.3962)	(0.0481)	(0.1102)	(0.0579)	(0.0823)
$ln(ConsumerLoans)\times(P_l)$	-0.1311***	-0.0942	-0.0922***	-0.0349***	-0.0450	-0.0271***	0.0089	-0.0053	-0.0823***
	(0.0248)	(0.0707)	(0.0182)	(0.0109)	(0.0823)	(0.0104)	(0.0142)	(0.0122)	(0.0136)
$ln(BusinessLoans)\times(P_l)$	0.0643***	0.0995	-0.0199	0.0283***	-0.0044	0.0868***	0.0423***	0.0476***	0.0509***
	(0.0091)	(0.0711)	(0.0185)	(0.0076)	(0.0272)	(0.0080)	(0.0068)	(0.0072)	(0.0058)
$ln(Securities)\times(P_l)$	-0.0261	0.2497**	0.0852***	-0.0242***	0.0349	-0.0505***	-0.0323***	-0.0958***	-0.0006
	(0.0242)	(0.1092)	(0.0238)	(0.0089)	(0.0780)	(0.0071)	(0.0101)	(0.0081)	(0.0108)
$ln(ConsumerLoans)\times(P_f)$	0.0100	0.0260	0.0267**	0.0021	-0.0540	0.0189***	0.0180*	0.0237***	-0.0052
	(0.0145)	(0.0459)	(0.0119)	(0.0063)	(0.0431)	(0.0072)	(0.0106)	(0.0075)	(0.0097)

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{CT})$ (10)	$\ln(TC_{DC})$ (11)	$\ln(TC_{DE})$ (12)	$\ln(TC_{FL})$ (13)	$\ln(TC_{HI})$ (14)	$\ln(TC_{IA})$ (15)	$\ln(TC_{IN})$ (16)	$\ln(TC_{KS})$ (17)	$\ln(TC_{KY})$ (18)
$ln(BusinessLoans)\times(P_f)$	-0.0364***	-0.0163	-0.0048	-0.0006	0.0116	-0.0445***	-0.0374***	-0.0280***	-0.0302***
	(0.0068)	(0.0448)	(0.0093)	(0.0043)	(0.0157)	(0.0060)	(0.0064)	(0.0050)	(0.0047)
$ln(Securities)\times(P_f)$	0.0370**	0.0505	0.0197	0.0269***	0.0857**	0.0367***	0.0085	0.0206***	0.0060
	(0.0159)	(0.0798)	(0.0175)	(0.0061)	(0.0394)	(0.0046)	(0.0072)	(0.0047)	(0.0083)
$ln(ConsumerLoans)\times(P_k)$	-0.0033	-0.0099	0.0281***	-0.0107***	0.0571*	0.0024	0.0010	0.0048	0.0051
	(0.0117)	(0.0240)	(0.0076)	(0.0041)	(0.0323)	(0.0046)	(0.0064)	(0.0053)	(0.0059)
$ln(BusinessLoans)\times(P_k)$	0.0160***	-0.0259	0.0063	-0.0148***	-0.0190*	-0.0166***	0.0109***	-0.0054*	-0.0078***
	(0.0051)	(0.0221)	(0.0086)	(0.0031)	(0.0108)	(0.0038)	(0.0039)	(0.0030)	(0.0030)
$ln(Securities)\times(P_k)$	-0.0186*	0.0060	-0.0401***	-0.0015	-0.0473	0.0144***	0.0043	0.0105***	-0.0045
	(0.0103)	(0.0383)	(0.0116)	(0.0034)	(0.0301)	(0.0023)	(0.0045)	(0.0034)	(0.0048)
$ln(P_l)\times(P_l)$	0.1248**	0.2843	0.0183	0.0881***	-0.0089	-0.0805***	0.0099	0.0037	-0.0403*
	(0.0525)	(0.1769)	(0.0708)	(0.0244)	(0.1389)	(0.0174)	(0.0297)	(0.0177)	(0.0208)
$ln(P_f)\times(P_f)$	0.0177	-0.1360	-0.0982***	0.0428***	0.0927***	0.0420***	0.1313***	0.0678***	-0.0120
	(0.0302)	(0.0996)	(0.0323)	(0.0117)	(0.0285)	(0.0088)	(0.0203)	(0.0100)	(0.0137)
$ln(P_l)\times(P_k)$	-0.0408	-0.0212	0.0878	-0.0583***	-0.0066	-0.0381***	0.0475*	-0.0452***	-0.0313*
	(0.0385)	(0.0985)	(0.0612)	(0.0170)	(0.0898)	(0.0110)	(0.0263)	(0.0128)	(0.0165)
$ln(P_k)\times ln(P_k)$	-0.0321**	0.0010	-0.0257*	-0.0063	0.0095	-0.0035**	0.0099	0.0029	-0.0042
	(0.0125)	(0.0275)	(0.0145)	(0.0042)	(0.0135)	(0.0015)	(0.0067)	(0.0032)	(0.0044)
$ln(P_l)\times(P_f)$	-0.2225***	0.1031	0.1179	-0.2121***	-0.2005*	-0.1053***	-0.2037***	-0.0679***	-0.0977***
	(0.0551)	(0.1569)	(0.0751)	(0.0256)	(0.1020)	(0.0163)	(0.0343)	(0.0193)	(0.0285)
$ln(P_f)\times(P_k)$	-0.0143	-0.0173	0.0234	0.0088	-0.1179***	0.0002	0.0407**	0.0305***	0.0107
	(0.0235)	(0.0553)	(0.0450)	(0.0098)	(0.0417)	(0.0076)	(0.0180)	(0.0086)	(0.0122)
$ln(LoanLossRation)$	-1.5667*	-3.8732	-0.0385	-1.0378	-4.5007**	-2.9080***	-1.5783***	-2.9302***	-3.4345***
	(0.8843)	(2.5001)	(0.5565)	(0.6499)	(2.0498)	(0.3954)	(0.5723)	(0.3645)	(0.4341)
$ln(LoanLossRation)^2$	43.2169***	-8.6980	0.1715	13.8975*	-0.7553	-8.0261	-7.6783	-4.6926	-8.8592***
	(13.8372)	(87.9357)	(0.1401)	(8.3585)	(19.4587)	(4.9496)	(8.4661)	(3.3090)	(1.5358)
$(TimeIndicator)$	-0.2359***	-0.0051	-0.1796*	-0.2909***	-0.0744	-0.0878***	-0.0884***	-0.1215***	-0.1030***
	(0.0278)	(0.0874)	(0.0948)	(0.0185)	(0.0566)	(0.0092)	(0.0161)	(0.0124)	(0.0135)
$(TimeIndicator)^2$	0.1582***	0.0567	0.1461***	0.2055***	0.0525**	0.0715***	0.0963***	0.0676***	0.0966***
	(0.0115)	(0.0358)	(0.0332)	(0.0071)	(0.0256)	(0.0035)	(0.0062)	(0.0046)	(0.0049)
Number of Observations	1,206	117	472	5,329	183	7,848	3,773	7,016	4,618
Number of banks	124	21	56	681	20	613	342	540	415
R-Squared	0.816	0.918	0.744	0.837	0.878	0.817	0.756	0.781	0.827

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{ME})$ (19)	$\ln(TC_{MN})$ (20)	$\ln(TC_{MS})$ (21)	$\ln(TC_{MT})$ (22)	$\ln(TC_{ND})$ (23)	$\ln(TC_{NE})$ (24)	$\ln(TC_{NJ})$ (25)	$\ln(TC_{NM})$ (26)	$\ln(TC_{NV})$ (27)
$ln(ConsumerLoans)$	0.3083**	-0.1721***	0.0468	0.2478**	0.1248	0.2761***	-0.0638	0.1369	0.5695***
	(0.1202)	(0.0518)	(0.1243)	(0.1145)	(0.0834)	(0.0508)	(0.0511)	(0.1513)	(0.1867)
$ln(BusinessLoans)$	-0.3954***	-0.4165***	-0.2873***	-0.3574***	-0.6201***	-0.3785***	-0.1945***	-0.2178***	0.0057
	(0.0940)	(0.0376)	(0.0571)	(0.0745)	(0.0784)	(0.0449)	(0.0257)	(0.0817)	(0.1428)
ln(Securities)	-0.2142**	0.1847***	0.8452***	0.2636***	0.3642***	0.3448***	0.2777***	0.5070***	0.4187**
	(0.1087)	(0.0386)	(0.0760)	(0.0768)	(0.0663)	(0.0355)	(0.0638)	(0.0919)	(0.1811)
$ln(ConsumerLoans)\times ln(ConsumerLoans)$	0.0230***	0.0181***	0.0255***	0.0014	0.0297***	0.0053***	0.0077***	0.0115	0.0056
	(0.0050)	(0.0025)	(0.0069)	(0.0065)	(0.0037)	(0.0018)	(0.0014)	(0.0071)	(0.0037)
$ln(BusinessLoans)\times ln(BusinessLoans)$	0.0103***	0.0260***	0.0130***	0.0161***	0.0445***	0.0234***	0.0108***	0.0172***	0.0115***
	(0.0015)	(0.0009)	(0.0012)	(0.0013)	(0.0038)	(0.0011)	(0.0007)	(0.0017)	(0.0041)
$ln(Securities)\times$ln(Securities)	0.0121***	0.0170***	0.0240***	0.0195***	0.0173***	0.0199***	0.0291***	0.0138***	0.0107***
	(0.0025)	(0.0007)	(0.0013)	(0.0013)	(0.0011)	(0.0007)	(0.0025)	(0.0015)	(0.0027)
$ln(ConsumerLoans)\times ln(BusinessLoans)$	-0.0027	0.0008	-0.0129**	0.0114**	-0.0330***	-0.0103***	-0.0061***	0.0070	-0.0162***
	(0.0040)	(0.0025)	(0.0053)	(0.0055)	(0.0054)	(0.0024)	(0.0011)	(0.0072)	(0.0044)
$ln(ConsumerLoans)\times$ln(Securities)	0.0064	0.0022	-0.0549***	0.0007	-0.0224***	0.0064**	-0.0143***	0.0002	-0.0098*
	(0.0058)	(0.0024)	(0.0058)	(0.0063)	(0.0040)	(0.0032)	(0.0028)	(0.0053)	(0.0054)
$ln(BusinessLoans)\times$ln(Securities)	-0.0060	-0.0120***	0.0088***	-0.0208***	0.0153***	-0.0095***	-0.0017	-0.0142***	-0.0393***
	(0.0041)	(0.0019)	(0.0023)	(0.0046)	(0.0040)	(0.0022)	(0.0015)	(0.0053)	(0.0064)
$ln(P_l)$	1.2791**	-0.8549***	1.0098***	-0.0844	-0.3322	0.7812***	1.4785***	1.0551**	1.2798
	(0.6338)	(0.1564)	(0.3375)	(0.3909)	(0.2876)	(0.1648)	(0.3167)	(0.4306)	(0.9564)
$ln(P_f)$	1.6845***	1.3173***	1.3763***	1.2232***	0.7679***	0.0741	1.2820***	2.0128***	0.6287
	(0.3752)	(0.0946)	(0.2183)	(0.2051)	(0.1374)	(0.0912)	(0.1918)	(0.2226)	(0.7068)
$ln(P_k)$	-0.6274**	0.5196***	0.2957***	1.0602***	0.2779***	0.2346***	0.1158	0.7688***	-0.7571*
	(0.2744)	(0.0543)	(0.1131)	(0.1339)	(0.1071)	(0.0577)	(0.1221)	(0.1864)	(0.4312)
$ln(ConsumerLoans)\times(P_l)$	-0.1050***	-0.0007	0.1033***	-0.0179	-0.0105	-0.0134	0.0296***	-0.1184***	-0.0479
	(0.0274)	(0.0098)	(0.0221)	(0.0244)	(0.0203)	(0.0108)	(0.0103)	(0.0294)	(0.0347)
$ln(BusinessLoans)\times(P_l)$	0.0455**	0.0394***	0.0346***	0.0814***	0.0590***	0.0536***	0.0250***	0.0518***	0.0932***
	(0.0218)	(0.0063)	(0.0118)	(0.0148)	(0.0205)	(0.0081)	(0.0055)	(0.0153)	(0.0286)
$ln(Securities)\times(P_l)$	-0.0053	-0.0165**	-0.1304***	-0.0543***	-0.0925***	-0.0767***	-0.0846***	-0.1292***	0.0133
	(0.0257)	(0.0082)	(0.0147)	(0.0147)	(0.0149)	(0.0083)	(0.0131)	(0.0226)	(0.0380)
$ln(ConsumerLoans)\times(P_f)$	0.0606***	0.0085	-0.0154*	0.0187	-0.0272**	0.0391***	-0.0206***	-0.0458***	0.0206
	(0.0183)	(0.0065)	(0.0089)	(0.0154)	(0.0107)	(0.0078)	(0.0063)	(0.0154)	(0.0136)

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{ME})$ (19)	$\ln(TC_{MN})$ (20)	$\ln(TC_{MS})$ (21)	$\ln(TC_{MT})$ (22)	$\ln(TC_{ND})$ (23)	$\ln(TC_{NE})$ (24)	$\ln(TC_{NJ})$ (25)	$\ln(TC_{NM})$ (26)	$\ln(TC_{NV})$ (27)
$ln(BusinessLoans)\times(P_f)$	-0.0664***	-0.0447***	-0.0297***	-0.0184*	0.0027	-0.0428***	-0.0241***	-0.0053	-0.0440**
	(0.0182)	(0.0052)	(0.0077)	(0.0111)	(0.0104)	(0.0065)	(0.0039)	(0.0093)	(0.0181)
$ln(Securities)\times(P_f)$	-0.0096	0.0508***	0.0642***	0.0153*	0.0228***	0.0469***	0.0414***	-0.0112	0.0421**
	(0.0196)	(0.0047)	(0.0094)	(0.0087)	(0.0048)	(0.0060)	(0.0087)	(0.0083)	(0.0213)
$ln(ConsumerLoans)\times(P_k)$	0.0720***	-0.0009	0.0310***	-0.0145	-0.0487***	0.0215***	-0.0308***	-0.0284**	-0.0010
	(0.0195)	(0.0046)	(0.0060)	(0.0127)	(0.0090)	(0.0048)	(0.0052)	(0.0131)	(0.0152)
$ln(BusinessLoans)\times(P_k)$	-0.0060	-0.0201***	0.0133***	0.0080	0.0025	-0.0283***	0.0062***	0.0128	-0.0203*
	(0.0103)	(0.0035)	(0.0039)	(0.0081)	(0.0082)	(0.0039)	(0.0024)	(0.0097)	(0.0121)
$ln(Securities)\times(P_k)$	-0.0361**	-0.0019	-0.0488***	0.0018	0.0003	0.0038	0.0179**	0.0127**	-0.0407**
	(0.0179)	(0.0033)	(0.0058)	(0.0077)	(0.0060)	(0.0028)	(0.0072)	(0.0059)	(0.0173)
$ln(P_l)\times(P_l)$	-0.1326**	0.0496***	-0.1879***	-0.0214	-0.0040	-0.0620***	-0.1246***	0.1098**	-0.2083*
	(0.0636)	(0.0161)	(0.0393)	(0.0412)	(0.0311)	(0.0177)	(0.0340)	(0.0475)	(0.1170)
$ln(P_f)\times(P_f)$	0.0878**	0.0620***	0.0313	0.0964***	-0.1007***	-0.0518***	0.0872***	0.1299***	0.0153
	(0.0353)	(0.0077)	(0.0194)	(0.0199)	(0.0086)	(0.0075)	(0.0185)	(0.0191)	(0.0605)
$ln(P_l)\times(P_k)$	-0.0264	-0.0799***	0.0273	-0.2269***	-0.0087	-0.0311***	-0.0324	-0.1899***	0.2287**
	(0.0636)	(0.0111)	(0.0235)	(0.0293)	(0.0249)	(0.0118)	(0.0255)	(0.0352)	(0.0910)
$ln(P_k)\times ln(P_k)$	-0.1390***	0.0161***	0.0074	0.0189***	0.0005	0.0052**	-0.0082	0.0237*	-0.0349
	(0.0353)	(0.0028)	(0.0053)	(0.0045)	(0.0066)	(0.0023)	(0.0075)	(0.0128)	(0.0240)
$ln(P_l)\times(P_f)$	-0.1078*	-0.1492***	-0.2324***	-0.0355	-0.2957***	-0.1054***	-0.0583*	-0.0206	-0.0937
	(0.0649)	(0.0160)	(0.0396)	(0.0367)	(0.0300)	(0.0193)	(0.0315)	(0.0367)	(0.1058)
$ln(P_f)\times(P_k)$	-0.0084	-0.0046	0.0940***	0.0648***	-0.0143	0.0375***	0.0097	0.0179	-0.1029**
	(0.0507)	(0.0069)	(0.0170)	(0.0153)	(0.0142)	(0.0076)	(0.0170)	(0.0228)	(0.0434)
$ln(LoanLossRation)$	2.1228	-4.8227***	-3.1297***	-1.6458**	-2.8838***	-1.4537***	-1.0863	-2.0024***	-1.0409
	(2.5676)	(0.4488)	(0.6055)	(0.7704)	(0.6097)	(0.4493)	(1.1459)	(0.6494)	(1.2185)
$ln(LoanLossRation)^2$	150.7171**	-27.4508***	-7.6330***	15.7818	19.1694**	-4.2692	18.7540	-8.7086***	-6.1739
	(75.6093)	(5.7564)	(2.3968)	(14.4272)	(8.4670)	(4.6428)	(17.1100)	(2.2264)	(7.8927)
$(TimeIndicator)$	-0.2118***	-0.1635***	-0.1755***	-0.1375***	-0.0497***	-0.0457***	-0.2186***	-0.1053***	-0.1953*
	(0.0295)	(0.0106)	(0.0199)	(0.0214)	(0.0160)	(0.0120)	(0.0227)	(0.0272)	(0.1005)
$(TimeIndicator)^2$	0.1784***	0.1001***	0.1125***	0.1059***	0.0413***	0.0424***	0.1724***	0.0880***	0.2348***
	(0.0117)	(0.0040)	(0.0075)	(0.0082)	(0.0061)	(0.0045)	(0.0088)	(0.0106)	(0.0366)
Number of Observations	671	8,714	1,909	1,605	1,963	5,133	2,397	1,101	448
Number of Banks	50	664	156	144	151	395	259	107	63
R-Squared	0.90	0.803	0.852	0.862	0.884	0.789	0.824	0.838	0.865

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{OR})$ (28)	$\ln(TC_{RI})$ (29)	$\ln(TC_{SD})$ (30)	$\ln(TC_{TX})$ (31)	$\ln(TC_{UT})$ (32)	$\ln(TC_{VA})$ 33)	$\ln(TC_{VT})$ (34)	$\ln(TC_{WV})$ (35)	$\ln(TC_{WY})$ (36)
$ln(ConsumerLoans)$	-0.0477	0.2491**	0.2381***	-0.0448	-0.0853	0.1910***	-0.3406**	0.0028	0.2009
	(0.1192)	(0.1014)	(0.0682)	(0.0336)	(0.0599)	(0.0600)	(0.1655)	(0.1439)	(0.1671)
$ln(BusinessLoans)$	-0.5140***	0.1828**	0.1058**	-0.4789***	0.1283**	-0.6076***	0.3779**	-0.5317***	-0.3230**
	(0.1068)	(0.0865)	(0.0500)	(0.0246)	(0.0629)	(0.0492)	(0.1856)	(0.0839)	(0.1359)
ln(Securities)	0.1836*	0.0029	-0.1172	0.2761***	0.1635**	0.0912	-0.0408	-0.1678*	0.2925**
	(0.1075)	(0.1212)	(0.0758)	(0.0201)	(0.0691)	(0.0575)	(0.1224)	(0.0933)	(0.1317)
$ln(ConsumerLoans)\times ln(ConsumerLoans)$	0.0085***	0.0132***	0.0128***	0.0355***	0.0413***	0.0126***	0.0252***	0.0165**	0.0363***
	(0.0023)	(0.0026)	(0.0028)	(0.0015)	(0.0022)	(0.0015)	(0.0069)	(0.0070)	(0.0088)
$ln(BusinessLoans)\times ln(BusinessLoans)$	0.0250***	0.0122***	0.0231***	0.0242***	0.0088***	0.0182***	0.0022	0.0106***	0.0199***
	(0.0022)	(0.0026)	(0.0019)	(0.0006)	(0.0014)	(0.0012)	(0.0031)	(0.0017)	(0.0026)
$ln(Securities)\times$ln(Securities)	0.0084***	0.0200***	0.0187***	0.0173***	0.0068***	0.0275***	-0.0030*	0.0213***	0.0245***
	(0.0022)	(0.0035)	(0.0021)	(0.0004)	(0.0015)	(0.0015)	(0.0016)	(0.0015)	(0.0024)
$ln(ConsumerLoans)\times ln(BusinessLoans)$	-0.0148***	-0.0143***	-0.0025	-0.0164***	-0.0187***	-0.0020	-0.0305***	0.0146**	-0.0520***
	(0.0040)	(0.0053)	(0.0018)	(0.0017)	(0.0024)	(0.0019)	(0.0112)	(0.0065)	(0.0106)
$ln(ConsumerLoans)\times$ln(Securities)	-0.0008	-0.0251***	-0.0175***	-0.0120***	-0.0053*	-0.0121***	0.0157**	-0.0310***	-0.0134
	(0.0044)	(0.0051)	(0.0033)	(0.0013)	(0.0027)	(0.0040)	(0.0071)	(0.0072)	(0.0111)
$ln(BusinessLoans)\times$ln(Securities)	0.0224***	-0.0162***	-0.0050**	-0.0067***	-0.0007	-0.0049**	-0.0035	0.0097**	-0.0064
	(0.0041)	(0.0042)	(0.0025)	(0.0011)	(0.0024)	(0.0023)	(0.0053)	(0.0044)	(0.0059)
$ln(P_l)$	1.8306**	-1.0506	0.2126	0.2741**	0.7016	-0.2025	-4.0240***	-3.7319***	-0.6583
	(0.7673)	(0.6389)	(0.4062)	(0.1316)	(0.4622)	(0.2003)	(1.0216)	(0.4005)	(0.6615)
$ln(P_f)$	0.7152*	-0.9427	0.9201***	1.3538***	-0.3393	1.3676***	1.4153***	1.8344***	0.7820**
	(0.3656)	(0.5919)	(0.2798)	(0.0753)	(0.3255)	(0.2349)	(0.4336)	(0.2723)	(0.3378)
$ln(P_k)$	-0.5124**	-0.0862	0.9861***	0.2008***	0.8285***	0.4891***	-0.7806**	1.2235***	0.5407***
	(0.2473)	(0.2371)	(0.1307)	(0.0414)	(0.1588)	(0.1225)	(0.3427)	(0.1998)	(0.1642)
$ln(ConsumerLoans)\times(P_l)$	0.0028	0.0243	-0.0357**	-0.0536***	-0.0335***	-0.0069	0.0634	-0.0127	0.0023
	(0.0284)	(0.0175)	(0.0166)	(0.0074)	(0.0112)	(0.0117)	(0.0463)	(0.0263)	(0.0363)
$ln(BusinessLoans)\times(P_l)$	0.0500**	0.0040	-0.0572***	0.1289***	0.0141	0.0617***	-0.0442	0.0688***	0.0884***
	(0.0245)	(0.0168)	(0.0134)	(0.0054)	(0.0122)	(0.0071)	(0.0528)	(0.0154)	(0.0293)
$ln(Securities)\times(P_l)$	-0.0913***	-0.0159	0.0459**	-0.0445***	-0.0259*	-0.0154	-0.0112	0.0418**	-0.0250
	(0.0213)	(0.0273)	(0.0194)	(0.0043)	(0.0140)	(0.0139)	(0.0305)	(0.0206)	(0.0258)
$ln(ConsumerLoans)\times(P_f)$	-0.0552***	0.0071	-0.0328***	-0.0278***	0.0093	0.0365***	0.0109	-0.0540***	-0.0019
	(0.0155)	(0.0111)	(0.0095)	(0.0043)	(0.0087)	(0.0118)	(0.0204)	(0.0163)	(0.0233)

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total cost of banks (continued from previous page)

	$\ln(TC_{OR})$ (28)	$\ln(TC_{RI})$ (29)	$\ln(TC_{SD})$ (30)	$\ln(TC_{TX})$ (31)	$\ln(TC_{UT})$ (32)	$\ln(TC_{VA})$ 33)	$\ln(TC_{VT})$ (34)	$\ln(TC_{WV})$ (35)	$\ln(TC_{WY})$ (36)
$ln(BusinessLoans) \times (P_f)$	0.0064	0.0175*	0.0235***	-0.0009	0.0081	-0.0864***	-0.0235	0.0086	-0.0843***
	(0.0146)	(0.0097)	(0.0071)	(0.0034)	(0.0091)	(0.0102)	(0.0255)	(0.0096)	(0.0206)
$ln(Securities) \times (P_f)$	0.0210*	-0.0179	0.0066	0.0267***	0.0128	0.0579***	-0.0020	0.0442***	0.0907***
	(0.0120)	(0.0166)	(0.0129)	(0.0024)	(0.0093)	(0.0122)	(0.0091)	(0.0116)	(0.0136)
$ln(ConsumerLoans) \times (P_k)$	0.0086	0.0218**	0.0165**	0.0175***	-0.0032	-0.0372***	0.0492***	0.0252	0.0247
	(0.0083)	(0.0092)	(0.0082)	(0.0024)	(0.0062)	(0.0073)	(0.0177)	(0.0160)	(0.0161)
$ln(BusinessLoans) \times (P_k)$	0.0231*	0.0386***	-0.0204***	-0.0244***	-0.0051	0.0011	-0.0040	-0.0299***	-0.0367***
	(0.0132)	(0.0120)	(0.0075)	(0.0023)	(0.0052)	(0.0044)	(0.0199)	(0.0085)	(0.0128)
$ln(Securities) \times (P_k)$	0.0045	-0.0497***	0.0014	-0.0052***	0.0127**	0.0117	-0.0019	-0.0768***	-0.0101
	(0.0085)	(0.0181)	(0.0058)	(0.0018)	(0.0054)	(0.0075)	(0.0092)	(0.0121)	(0.0113)
$ln(P_l) \times (P_l)$	-0.1610*	0.2566***	0.0310	-0.0639***	0.0731*	-0.0540***	0.5190***	0.3138***	0.0654
	(0.0872)	(0.0835)	(0.0427)	(0.0136)	(0.0438)	(0.0130)	(0.1561)	(0.0409)	(0.0813)
$ln(P_f) \times (P_f)$	0.0572*	-0.0729	0.0737***	0.1048***	0.0291	0.0440**	0.0859**	0.1080***	0.1036***
	(0.0308)	(0.0704)	(0.0282)	(0.0068)	(0.0313)	(0.0214)	(0.0383)	(0.0246)	(0.0329)
$ln(P_l) \times (P_k)$	-0.0775	-0.0093	-0.1973***	-0.0534***	-0.1978***	-0.0502**	0.1005	-0.1578***	-0.1474***
	(0.0598)	(0.0574)	(0.0313)	(0.0098)	(0.0335)	(0.0245)	(0.1097)	(0.0415)	(0.0416)
$ln(P_k) \times ln(P_k)$	-0.0319***	0.0307	0.0054	-0.0064***	0.0179**	-0.0256**	0.0049	-0.0277**	-0.0485***
	(0.0117)	(0.0239)	(0.0087)	(0.0020)	(0.0083)	(0.0106)	(0.0362)	(0.0131)	(0.0090)
$ln(P_l) \times (P_f)$	0.0350	0.1859	0.0206	-0.0648***	0.1806**	-0.1565***	-0.0872	-0.1806***	0.0614
	(0.0638)	(0.1159)	(0.0440)	(0.0128)	(0.0758)	(0.0374)	(0.1283)	(0.0437)	(0.0648)
$ln(P_f) \times (P_k)$	-0.1168***	-0.0334	0.0644***	-0.0196***	0.0223	0.0610***	-0.0052	0.0008	-0.0101
	(0.0328)	(0.0545)	(0.0227)	(0.0058)	(0.0251)	(0.0213)	(0.0498)	(0.0261)	(0.0230)
$ln(LoanLossRation)$	0.1024	-1.3032	-5.8338***	-1.5183***	-1.6961***	-3.2878***	-4.3527	-0.2121	-4.0049***
	(1.1635)	(3.1537)	(0.6252)	(0.2772)	(0.4926)	(0.7443)	(3.1782)	(1.0612)	(1.3364)
$ln(LoanLossRation)^2$	1.2338	-57.4343	-11.0991***	0.9440	1.6450***	-3.7747***	48.4476	17.0348	-21.8816***
	(4.0489)	(64.0906)	(2.7973)	(2.8508)	(0.5546)	(1.4460)	(98.0243)	(13.5533)	(8.4191)
$(TimeIndicator)$	-0.2619***	-0.0164	-0.0578**	-0.1621***	-0.1968***	-0.3084***	-0.1543***	-0.0694***	-0.1441***
	(0.0472)	(0.0542)	(0.0274)	(0.0094)	(0.0632)	(0.0215)	(0.0368)	(0.0194)	(0.0372)
$(TimeIndicator)^2$	0.2372***	0.0984***	0.0545***	0.0911***	0.1865***	0.1929***	0.1463***	0.0847***	0.0866***
	(0.0194)	(0.0198)	(0.0104)	(0.0036)	(0.0227)	(0.0084)	(0.0155)	(0.0076)	(0.0149)
Number of Observations	701	208	1,768	13,689	704	2,626	388	1,591	872
Number of Banks	83	22	135	1203	76	285	30	175	69
R-Squared	0.885	0.917	0.752	0.807	0.881	0.833	0.806	0.831	0.826

Notes: Dependent variable: logarithm of total cost. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 1.3: The effect of contributions on the Lerner index

	(1) Lerner Index	(2) Lerner Index	(3) Lerner Index
Contributions	0.0502	0.0399	0.0412
(millions of dollars)	(0.0307)	(0.0268)	(0.0286)
Lag of Contributions	0.0698**	0.0541**	0.0564**
(millions of dollars)	(0.0288)	(0.0244)	(0.0254)
Asset Size	0.0017***	0.0011***	0.0012***
(billions of dollars)	(0.0005)	(0.0004)	(0.0004)
ROA	0.0322***	0.0326***	0.0357***
	(0.0011)	(0.0010)	(0.0011)
Number of Employees	-0.6905***	-0.5933***	-0.6120***
(10,000's)	(0.2595)	(0.1808)	(0.1898)
Number of Offices	-0.0273***	-0.0177**	-0.0185**
(100's)	(0.0095)	(0.0073)	(0.0074)
State dummies (For TC)	N	Y	N
Total Cost Separately	N	N	Y
Number of Observations	96,224	96,224	96,224
Number of Banks	8,837	8,837	8,837
R-Squared	0.347	0.392	0.397

Notes: Dependent variable: the value of the Lerner index. All regressions include following dummy variables: a dummy equal to one in election years, a dummy equal to one for the year in which a structural change was implemented by the institution, a dummy equal to one if the institution can operate in more than one state, and a full set of institution's chart class, asset concentration specialization and insurance fund membership, bank, state and time dummies. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 1.4: The effect of contributions on the Lerner index (2 lags)

	(1) Lerner Index	(2) Lerner Index	(3) Lerner Index
Contributions	0.0478	0.0380	0.0392
(millions of dollars)	(0.0299)	(0.0262)	(0.0281)
Lag of Contributions	0.0731**	0.0567**	0.0590**
(millions of dollars)	(0.0310)	(0.0262)	(0.0273)
Second Lag of Contributions	0.0434	0.0332	0.0343
(millions of dollars)	(0.0289)	(0.0265)	(0.0276)
Asset Size	0.0017***	0.0011***	0.0012***
(billions of dollars)	(0.0005)	(0.0004)	(0.0004)
ROA	0.0322***	0.0326***	0.0357***
	(0.0011)	(0.0010)	(0.0011)
Number of Employees	-0.6912***	-0.5939***	-0.6125***
(10,000's)	(0.2596)	(0.1808)	(0.1899)
Number of Offices	-0.0275***	-0.0178**	-0.0186**
(100's)	(0.0095)	(0.0074)	(0.0075)
State dummies (For TC)	N	Y	N
Total Cost Separately	N	N	Y
Number of Observations	96,224	96,224	96,224
Number of Banks	8,837	8,837	8,837
R-Squared	0.349	0.392	0.397

Notes: Dependent variable: the value of the Lerner index. All regressions include following dummy variables: a dummy equal to one in election years, a dummy equal to one for the year in which a structural change was implemented by the institution, a dummy equal to one if the institution can operate in more than one state, and a full set of institution's chart class, asset concentration specialization and insurance fund membership, bank, state and time dummies. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 1.5: The effect of contributions on H-Statistics

	(1) H-Statistic	(2) H-Statistic
Contributions	-0.0257	-0.0255
(millions of dollars)	(0.0160)	(0.0162)
Lag of Contributions	-0.0490***	-0.0489***
(millions of dollars)	(0.0142)	(0.0144)
Asset Size	-0.0004	-0.0004
(billions of dollars)	(0.0002)	(0.0003)
ROA	0.0015***	0.0014***
	(0.0004)	(0.0004)
Number of Employees	0.0566	0.0552
(10,000's)	(0.1340)	(0.1359)
Number of Offices	-0.0023	-0.0024
(100's)	(0.0023)	(0.0024)
State dummies (For TC)	N	Y
Number of Observations	96,224	96,224
Number of Groups	8,837	8,837
R-Squared	0.801	0.795

Notes: Dependent variable: the value of H-Statistics. All regressions include following dummy variables: a dummy equal to one in election years, a dummy equal to one for the year in which a structural change was implemented by the institution, a dummy equal to one if the institution can operate in more than one state, and a full set of institution's chart class, asset concentration specialization and insurance fund membership, bank, state and time dummies. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 1.6: Estimation of total revenue of banks

	ln(TR)	ln(TR)
	(1)	(2)
ln(ConsumerLoans)	0.0816***	0.0811***
	-0.0117	(0.0117)
ln(BusinessLoans)	-0.1191***	-0.1213***
	-0.0086	(0.0086)
ln(Securities)	0.0290***	0.0253**
	-0.0101	(0.0101)
ln(LoanLossRatio)	-2.2605***	-2.0720***
	-0.6824	(0.6825)
ln(DepositetoLiabilities)	-0.8731***	-0.8789***
	-0.0639	(0.0639)
ln(ConsumerLoans)×*ln(ConsumerLoans)*	0.0198***	0.0199***
	-0.0003	(0.0003)
ln(BusinessLoans)×*ln(BusinessLoans)*	0.0211***	0.0210***
	-0.0002	(0.0002)
ln(Securities)×*ln(Securities)*	0.0185***	0.0185***
	-0.0002	(0.0002)
ln(LoanLossRatio)×*ln(LoanLossRatio)*	0.2614***	0.2746***
	-0.0486	(0.0486)
ln(DepositetoLiabilities)×*ln(DepositetoLiabilities)*	-0.0291***	-0.0284***
	-0.0027	(0.0027)
ln(ConsumerLoans)×*ln(BusinessLoans)*	-0.0101***	-0.0101***
	-0.0003	(0.0003)
ln(ConsumerLoans)×*ln(Securities)*	-0.0112***	-0.0112***
	-0.0004	(0.0004)
ln(ConsumerLoans)×*ln(LoanLossRatio)*	-0.3059***	-0.3132***
	-0.0237	(0.0237)
ln(ConsumerLoans)×*ln(DepositetoLiabilities)*	0.0180***	0.0188***
	-0.0026	(0.0026)
ln(BusinessLoans)×*ln(Securities)*	-0.0069***	-0.0068***
	-0.0003	(0.0003)
ln(BusinessLoans)×*ln(LoanLossRatio)*	0.1112***	0.1093***
	-0.0264	(0.0264)
ln(BusinessLoans)×*ln(DepositetoLiabilities)*	0.0097***	0.0098***
	-0.0012	(0.0012)
ln(Securities)×*ln(LoanLossRatio)*	-0.1146***	-0.1121***
	-0.0285	(0.0285)
ln(Securities)×*ln(DepositetoLiabilities)*	0.0005	0.0005
	-0.0023	(0.0023)

Dependent variable: logarithm of total revenue. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total revenue of banks (continued from previous page)

	ln(TR)	ln(TR)
	(1)	(2)
ln(LoanLossRatio)×ln(DepositetoLiabilities)	-0.7015***	-0.6961***
	-0.1109	(0.1109)
$ln(P_l)$	0.1967***	0.1939***
	-0.0505	(0.0506)
$ln(P_f)$	0.8375***	0.8359***
	-0.0344	(0.0344)
$ln(P_k)$	0.1645***	0.1698***
	-0.0201	(0.0201)
$ln(ConsumerLoans)\times ln(P_l)$	-0.0106***	-0.0105***
	-0.0024	(0.0024)
$ln(BusinessLoans)\times ln(P_l)$	0.0168***	0.0169***
	-0.0017	(0.0017)
$ln(Securities)\times ln(P_l)$	-0.0045**	-0.0044**
	-0.0022	(0.0022)
$ln(LoanLossRatio)\times ln(P_l)$	0.8116***	0.7916***
	-0.1375	(0.1373)
$ln(DepositetoLiabilities)\times ln(P_l)$	0.0946***	0.0952***
	-0.0117	(0.0117)
$ln(ConsumerLoans)\times ln(P_f)$	-0.0003	0.0002
	-0.0015	(0.0015)
$ln(BusinessLoans)\times ln(P_f)$	-0.0052***	-0.0055***
	-0.0012	(0.0012)
$ln(Securities)\times ln(P_f)$	-0.0095***	-0.0102***
	-0.0013	(0.0013)
$ln(LoanLossRatio)\times ln(P_f)$	-0.4968***	-0.4659***
	-0.0879	(0.0880)
$ln(DepositetoLiabilities)\times ln(P_f)$	-0.0170**	-0.0164**
	-0.0074	(0.0074)
$ln(ConsumerLoans)\times ln(P_k)$	0.0017*	0.0017*
	-0.0009	(0.0009)
$ln(BusinessLoans)\times ln(P_k)$	-0.0038***	-0.0035***
	-0.0007	(0.0007)
$ln(Securities)\times ln(P_k)$	-0.0065***	-0.0068***
	-0.0008	(0.0008)
$ln(LoanLossRatio)\times ln(P_k)$	-0.3918***	-0.3990***
	-0.0667	(0.0667)

Notes: Dependent variable: logarithm of total revenue. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

continued on next page

Estimation of total revenue of banks (continued from previous page)

	ln(TR)	ln(TR)
	(1)	(2)
$ln(DepositetoLiabilities) \times ln(P_k)$	-0.0406***	-0.0408***
	-0.0053	(0.0053)
$ln(P_l) \times ln(P_l)$	-0.0259***	-0.0255***
	-0.0048	(0.0048)
$ln(P_f) \times ln(P_f)$	0.0005	0.0003
	-0.0029	(0.0029)
ln(Average Cost of Capital)$\times ln(P_k)$	0.0026***	0.0025***
	-0.0008	(0.0008)
$ln(P_l) \times ln(P_k)$	-0.0308***	-0.0316***
	-0.0043	(0.0043)
$ln(P_l) \times ln(P_f)$	-0.0738***	-0.0721***
	-0.0068	(0.0068)
$ln(P_f) \times ln(P_k)$	-0.0188***	-0.0181***
	-0.0028	(0.0028)
$Timeindicator$	-0.0121***	-0.0123***
	-0.0044	(0.0044)
$(Timeindicator)^2$	0	-0.0000
	-0.0001	(0.0001)
$(Timeindicator) \times ln(ConsumerLoans)$	-0.0004**	-0.0004**
	-0.0002	(0.0002)
$(Timeindicator) \times ln(BusinessLoans)$	0.0042***	0.0043***
	-0.0001	(0.0001)
$(Timeindicator) \times ln(Securities)$	-0.0051***	-0.0051***
	-0.0002	(0.0002)
$(Timeindicator) \times ln(LoanLossRatio)$	-0.1164***	-0.1121***
	-0.015	(0.0150)
$(Timeindicator) \times ln(DepositetoLiabilities)$	-0.0150***	-0.0150***
	-0.0013	(0.0013)
$(Timeindicator) \times ln(P_l)$	0.0043***	0.0044***
	-0.0009	(0.0009)
$(Timeindicator) \times ln(P_k)$	-0.0046***	-0.0045***
	-0.0003	(0.0003)
$(Timeindicator) \times ln(P_f)$	-0.0090***	-0.0091***
	-0.0007	(0.0007)
State dummies	N	Y
Number of Observations	96,224	96,224
Number of banks	8,837	8,837
R-Squared	0.779	0.779

Notes: Dependent variable: logarithm of total revenue. All regressions include a full set of institution dummies. P_{Ft}, is the average interest rate paid for all domestic and foreign deposits and liability funding; P_{Lt}, is the annual average expense per full-time equivalent worker; and P_{Kt}, an approximation to the cost of capital. Clustered standard errors (at the bank level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Chapter 2

Market Structure and the Electoral Power of Special Interest Groups

2.1 Introduction

Campaign contributions from special interest groups (SIGs) have become one of the major sources of funding for political campaigns. Theoretical and empirical literatures have shown that contributions have a positive effect on the percentage of votes received by a candidate (Snyder, 1990, 1993; Grossman and Helpman, 1996). This positive effect can be explained by considering two sides of lobbying markets. On the one hand, politicians use these campaign contributions for advertising and informative activities to maximize their likelihood of winning an office by persuading voters (see, e.g., Baron, 1994). On the other hand, SIGs enforce social norms among their members by encouraging them to vote for the candidate with the most closely-aligned interests (see, e.g., Morton, 1987, 1991; Grossman and Helpman, 2001). This social norm enforcement paradigm can be applied to firms since they can be considered special interest groups. This approach leads to a natural question: can a bigger firm be more efficient in enforcing social norms and affecting the results of elections?

In this paper, I consider firms as special interest groups, and investigate the effect of market power on firms' electoral influence. Similar to the paper by Snyder (1990), given the contribution limits on individual contributions[1] and the small amount of contributions made by each firm, I assume that profit-seeking contributors are not looking for a change in major public policies, and I restrict my attention to the electoral power of interest groups. Based on the collective action theory proposed by Olson (1965), firms in more concentrated industries have more incentives to overcome the free-rider problem and behave as an organized interest group to advance their interests. To test the hypothesis that concentrated firms have more electoral power, this paper uses a new detailed dataset of the elections for the U.S. House of Representatives from 1990 to 2014, which includes all individual contributions made to Political Action Committees (PACs) or candidates directly, organized by donor occupation. This paper just includes the contributions made by profit-seeking firms and categorizes the contributions based on the concentration ratio of the industry. The concentration ratio of the industry is defined as the percentage of output accounted for by the largest 50 firms of total output of the industry. The related data are obtained from the U.S. Economic Census and assigned to the contributions based on their industry codes.

The results confirm that, even when controlling for the total value of contributions made to a candidate, political donations made by firms with high market power have a positive effect on the probability of winning. Within a race, contributions financed by more concentrated firms explain 29.54 percent of the variation in the probability of winning. This is in addition to the 25.42 percent accounted for by the total amount of contributions.[2]

Snyder (1990) argues that candidates with a high probability of winning (more than 85%) might accept less in contributions. As a robustness check, I run a subsample analysis on the close open-seat races[3] and find that the results of the paper are insensitive to the closeness of the races. The results are also robust to different ways of clustering the error terms.

[1]For more information visit: `http://www.fec.gov/pages/brochures/fecfeca.shtml#Contribution_Limits`.

[2]The results suggest that for a one standard deviation increase in the total contributions received by a candidate the likelihood of being elected increases by 27.97 percent.

[3]In which in which the vote share of the candidates are between 21.71% and 75.55%.

The rest of the paper proceeds as follows. Section 2 provides a survey of previous literature on the political influence of special interest groups. In Section 3, I build on the voting turnout literature to introduce the mechanisms that link the economic power of special interest groups to the behavior of voters and election results. Sections 4 and 5 discuss data sources and the empirical model of the paper, respectively. Section 6 presents the results of the empirical analysis and a series of robustness checks, and the last section discusses the conclusions of the paper.

2.2 Literature review

A number of studies have investigated the effect of special interest groups on elections. Typically, interest groups are assumed to have substantial political power to either change the candidates' position in their favor (influence channel)[4] or to promote their favorite candidates (electoral channel).[5] In this study I am specifically interested in the latter channel of contributions. A large number of studies have investigated the effect of campaign contributions on the electoral outcome of elections. Although they do not deny the existence of the influence channel which implies that campaign contributions are given in return for some private benefits, they argue that this effect is negligible (e.g., Tullock, 1972; Morton, 1991; Snyder, 1990, 1993; Baron, 1994).[6] These papers show that campaign spending has a positive effect on the percentage of votes received by a candidate and hence the chances of winning an election.

In the model presented by Baron (1994) voters are divided into two groups: informed

[4] A well-developed literature has investigated the influence channel and shown that the donations made by profit-seeking interest groups are considered quid pro quo contributions for specific favors. First, there are studies that explore the effect of contributions on a specific policy (e.g., Grossman and Helpman, 1994; Goldberg and Maggi, 1999; Mitra et al., 2002; Mayda et al., 2010; Igan and Mishra, 2011; Perez Saiz and Semenov, 2015). Second, there are studies that investigate the influence of contributions on firms' performance indices (e.g., Claessens et al., 2008; Bertrand et al., 2004; Gordon et al., 2007; Kim, 2008; Faccio, 2006; Igan et al., 2011; Chen et al., 2015).

[5] Grossman and Helpman (1996).

[6] In the period of 1907-1966, in response to President Roosevelt's concern regarding to the corporate contributions received by candidates, Congress of the U.S. passed several statutes in order to limit the influence of rich individuals and organizations on the federal election outcomes.

voters who vote based on the candidates' pronounced policies and uninformed voters who are under the influence of candidates' campaign advertising. He argues that catering the policies to the interest of a special group might cost candidates some informed voters. However, he shows that the productivity of campaign contributions increases with the number of uninformed voters. Therefore, as the number of uninformed voters increases, candidates try to raise more money to compete for uninformed voters. The empirical results of the model support the latter idea and show that probability of winning of a candidate is equal to his share of total investor-contributions.

In the papers by Snyder (1990, 1993), contributions are defined as a kind of asset market where the returns on quid pro quo contributions (investment contributions) are contingent on the election outcome. In the models, contributors are categorized into two main groups (Investors and Ideological). Snyder argues that investor-contributions do not have major effects on the policy outcomes of the elections[7] and shows that, in contrast to lopsided elections, candidates in close-races try to maximize the contributions by selling as many favors as they can. Using data on the U.S. Senate and House elections between 1980-1986, he shows that empirical results support the predictions of the model very well. He suggests that although the total amount is important, future studies should consider investigating the specific effect of different components of campaign contributions on election outcomes.

Morton (1991) argues that if candidates only care about the electoral outcome of the election, they would all cater to the median of the voters and a zero turnout could not be rejected. Therefore, a positive turnout on an election can only be explained by mixed motives (policies and winning) of candidates. The paper assumes that candidates hold distinct fixed positions and by including campaign contributions in the utility function of the voters, suggest that in an election with a large number of voters, political activities of interest groups provide a rational explanation for election turnout, and argues that groups' characteristics affect election outcomes.

[7]This is consistent with the model presented by Tullock (1972).

The only paper that considers the effect of different sources of campaign contributions is by Depken (1998). This paper investigates the effect of three sources of campaign contributions: political parties, PACs, and individual contributions on the probability of the winning of candidates, using the data for U.S. House and Senate elections in 1996. The results show that in both types of elections, PACs and individual contributions have a positive and significant effect on the probability of the winning, while the effect of party contributions is insignificant.

A few studies cast light on the impact of market structure on the influence power of interest groups. For example, Esty and Caves (1983), using data from thirty-five manufacturing industries in the US and different measures of political influence, show that the political influence of firms increases with their concentration and geopolitical dispersion. Pittman (1977) using the data of individuals who contributed more than 5,000 dollars in total between 1971 and 1972 in the U.S. presidential race, shows a positive relationship between the concentration ratio and rent-seeking political activities of the firms. The paper by Salamon and Siegfried (1977), using the data of 110 minor industries in manufacturing and mining, shows that larger firms are better in using their political power to avoid both federal corporate income and state excise taxes. The results that industries with more firms have less political power is consistent with the hypothesis of free riding in large industries. The paper by Macher et al. (2011) uses the data of 6,000 international firms to evaluate the influence of firms on changing government policies. The results suggest that firms' characteristics affect their influence on governments. For example, bigger firms or firms with fewer number of competitors are more efficient in affecting government policies.

2.3 The mechanism of influence

To explain the mechanism of the special interest groups' influence on individual voting behavior, I begin by explaining the rational behavior of voters. The question of why individuals participate in election has created a debate between sociologists and economists. A rational individual votes if the benefit of voting outweighs its cost. When a significant number of citizens are eligible to vote, the benefit of voting becomes negligible, so it is expected that a rational individual would chose not to spend time and effort to learn about the candidates and visit a polling station. Considering the large number of eligible voters in the United States, it is surprising to learn that the turnout in congressional elections is high, close to 50 percent[8], for the past two decades (the "paradox of voting" suggested by Downs, 1975).

A wide range of economic literature has tried to approach the motivation of the voters from different angles.[9] For many, the "social norms" hypothesis, based on the theory by Morton (1987, 1991), successfully solves the paradox of voting and explains the variation in the turnout rates between different societies.[10] Following the social norm theory of voting, the results of the paper by Blais (2000) suggest that voting is considered a citizen's duty and individuals participate in elections because they are afraid of being abandoned by their society.

Downs (1957) explains that social norms, in contrast to other kinds of norms (moral, legal and private), are enforced by members of a community and argues that when a certain behavior becomes a norm in a society, individuals have enough incentive to obey the norm in order to prevent any sanctions from other members of the group. Grossman and Helpman (2001) argue that in regard to participation in political voting, special interest groups matter a great deal for enforcing social norms. They define a special interest group as a set of individuals who share similar characteristics and benefit from policies that might not be in line with the interest of the average citizen. In this paper I consider individuals who work in the same industry as members of an interest group, as they share the same interest with regard to policies and regulations that affect their career.[11]

[8]File (2015).

[9]See Grossman and Helpman (2001) for a complete survey.

[10]A well developed literature has shown that social norms govern behavior of voters in groups and societies (e.g., Elster, 1989; Uhlaner, 1989; Shachar and Nalebuff, 1999; Bönisch et al., 2015; Campbell et al., 1954; Downs, 1957; Riker and Ordeshook, 1968; Knack, 1992).

[11]This is similar to the assumption of the paper by Grossman and Helpman (1994).

According to Baron (1994), gathering information about candidates' policy stances is an expensive and time-consuming procedure, so some individuals might not find it worthwhile to spend time and effort learning about the position of a candidate. He categorizes citizens into two groups: "informed voters" who are well-informed about candidates' prospects and policies and will chose the candidate who is more likely to satisfy their desires; and there are "uninformed voters" who are not fully aware of the situation and can be educated either by SIGs or candidates' electoral advertising which are funded by campaign contributions. Based on Grossman and Helpman (1994), contributors can be considered informed members of interest groups who contribute to either increase the probability of their desired candidates winning or to affect the pronounced policy of the candidates. According to Grossman and Helpman (2001), in both cases SIGs have an incentive to educate their uninformed members: first, because policy platforms cater to the interests of groups with more informed members (influence incentive); and second, because educated members are more likely to turn up on the election (electoral incentive).

Profit-seeking firms engage in political activities when they benefit from having a specific candidate in power. Although it is well documented that contributions from profit seeking firms are quid pro quo, I assume that investor contributors seek small favors such as tax exemptions or priority consideration when the government chooses a supplier[12] and focus on the elector impact of SIGs. Firms conduct research and gather information on policies that are related to the interests of their members. They use the information to persuade the candidate to change their policy platform in their favor and to educate their members and encourage them to vote for the candidate with whom their interests are most closely aligned. Firms are able to monitor the behavior of their employees through frequent social interactions and successfully abandon those who fail to obey (Grossman and Helpman, 2001). It is unlikely that SIGs encourage the participation of those who are likely to vote for the rivals of their desired candidate. Therefore, the results of election are contingent to the

[12]This is similar to papers by Tullock (1972), Baron (1994), Snyder (1990, 1993).

turnout of the members of SIGs.

The previous literature demonstrates a positive link between economic and political power. According to the collective action theory presented by Olson (1965), firms with fewer competitors are less likely to suffer from free-rider problems and are able to behave as an organized group to pursue their interests. To see the effect of the concentration ratio on the electoral power of the interest groups, controlling for the total amount of investor-contributions received by a candidate, I categorize the investment-contributions made by firms' employees based on the concentration ratio of the contributors' industries.

2.3.1 Testable hypotheses

Based on the mechanism of influence of special interest groups, this paper examines the following hypotheses:

1. Contributions have a positive impact on a candidate's probability of winning.

2. Firms in a concentrated market have more influence on election outcomes than firms with many rivals.

2.4 Data

In order to investigate the industry-specific effect of contributions on the probability of the winning an election, I use data on House of Representative elections in the United States between 1990 and 2014. I use U.S. Economic Census data to identify the concentration ratio associated with contributions. I also use the yearly Census of County Business Patterns in a test of external validity.

2.4.1 Campaign contributions data

Candidates running for federal, state, or local offices finance their electoral campaigns' activities through campaign contributions. For federal level elections (the office of the U.S. president, vice president, Senate, or House of Representative seats), campaign contributions are governed by federal law and enforced by the Federal Election Commission (FEC).[13] The Federal Election Campaign Act (FECA) requires candidates, parties and PACs to file and report any contributions or expenses that exceed 200 dollars per election cycle. Table 2.3 shows the limits imposed by the FECA on individuals and groups contributions made to candidates per election cycle.

[insert Table 2.3 here]

Contributions received by candidates are obtained from data collected by the Center for Responsive Politics (CRP).[14] The CRP is a non-partisan, independent nonprofit which provides the most comprehensive resource for federal campaign contributions and lobbying data.[15] It is funded by a combination of institutional grants, individual contributions and income earned from custom commercial researches.

Each person who runs for a political position gets a unique Candidate Identifier Code (CID) assigned by Federal Election Commission (FEC) which is constant throughout all cycles. I use the CID to identify the candidates, calculate the total contributions they receive, and their incumbency status. The dataset includes contributions made by individuals to candidates and to PACs, contributions from PACs to candidates and between PACs. Contributions have been linked to an industry code based on the occupation of donors. In order to trace the money flow accurately, if an individual contributes to a PAC instead of the candidate, the record of contribution appears twice, once when the donor gives money to the PAC and second, when the PAC gives money to the candidate. In the latter case, the

[13] Federal Election Commission (2014).

[14] Data are available at `https://www.opensecrets.org`.

[15] The CRP won the Society of Professional Journalists' 2013 award for public service in online journalism.

record is linked to an industry code based on the actual source of the contribution. To cover all the contributions received by candidates, I combine PAC and individual contributions. I exclude the individual contributions to PACs to avoid double counting them.

CRP's industry codes are compromised five characters. The industry codes have been categorized into 13 industrial groups and three non-industrial categories,[16] and at the detailed level there are more than 400 subgroups. To investigate the effect of industry contributions, I exclude the contributions made by non-industrial groups (i.e., contributions under Other, Unknown, and Candidate Committee have been excluded). In total, this paper uses the contributions made by 285 industries between 1990 and 2014.

2.4.2 Concentration ratio data

Concentration ratio data are obtained from the Economic Census of the United States.[17] The census is conducted by the U.S. Census Bureau every five years and based on federal law, all firms are required to answer. For the purpose of this study, I use the average of "the percentage share of the 50 largest firms of total sales, receipts, or revenue" as a measure of concentration over the 1997-2012 period. The concentration data is not available for the construction industry. In this case, I use "the percentage share of the 50 firms with the highest number of employees of the net value of business done" as an alternative measure of concentration.

This study uses data from four censuses in 1997, 2002, 2007, and 2012. In the censuses prior to 1997 the concentration ratios are reported based on a different industry classification coding and detailed data are missing for several industries. However, it is worthwhile to mention that the variation of the concentration ratio within an industry over the 15 years is less than 10 percent. In addition, as a robustness check, I run the regression using the exact

[16] Including: 1) Agriculture 2) Construction and Public Works 3) Communication and Electronics 4) Defence 5) Energy, Natural Resources and Environment 6) Finance, Insurance and Real Estate 7) General Commerce 8) Health, Education and Human Resources 9) Ideological and Single Issue 10) Legal Services 11) Labor Unions 12) Manufacturing 13) Transportation 14) Other 15) Unknown 16) Candidate Committees.

[17] Data are available at: `http://www.census.gov/econ/census/help/sector/data_topics/concentration_ratios.html`.

value of the concentration ratios for the 1997-2014 period.

While the U.S. Economic Census collects and reports the information based on the North American Industry Classification System (NAICS),[18] CRP has a different method for classifying industries. The basic industry code mapping from CRP to NAICS was made publicly available by David Kraljic,[19] the co-founder of the VoteTocracy website, for which I provided some updates.[20] I use the mapping between CRP and NAICS codes to link contributions to concentration ratios. To see the effect of different sources of campaign contributions on the candidates' probability of the winning, I categorized contributions into 10 different groups based on the level of concentration ratio of the industry. For example, Category 1 represents the industries with a concentration ratio between 0 and 10 (which includes contributions made by firms in a competitive market) and the last group is associated with the concentration ratio between 90 and 100 (firms that operate in a concentrated market).

2.4.3 Election results data

The election results data are obtained from the FEC.[21] The FEC collects the certified federal election results from state election offices.

I restrict attention to open-seat races (i.e., in which no incumbents ran).[22] Snyder (1990) argues that candidates try to raise more money by promising favors to contributors. He argues that incumbents are able to deliver more favors as they are more experienced and benefit from their seniority position, and are therefore, more likely to receive money.[23] I avoid this complication by only including open-seat races. I consider the races in which only

[18] NAICS is the federally approved standard way of identifying and collecting and information of the different U.S. industries. Related data can be reached at `http://www.census.gov/eos/www/naics`.

[19] Available at: `https://groups.google.com/forum/#!topic/opensecrets-open-data/nXYSeFrtwxk`.

[20] Coding available upon request. In order to match the codes accurately, I use the information about the detailed way of coding the contributions available at `http://www.opensecrets.org/industries/slist.php` and the list of industries can be found at `https://www.opensecrets.org/downloads/crp/CRP_Categories.txt`.

[21] Data are available at: `http://www.fec.gov/pubrec/electionresults.shtml`.

[22] There have been 600 open races between 1990 and 2014.

[23] There were 4920 races with an incumbent over the mentioned period and in more than 95 percent of the time, the incumbent re-won the office.

one Democrat and one Republican candidate ran in the general election.[24] Candidates who had been defeated in primary elections have been excluded. I also exclude the races in which the sum of third parties' vote share exceeded 10 percent.[25]

Some economists argue that incumbency privilege can go beyond the incumbency of the candidate.[26] They argue that candidates from the incumbent party might benefit from positive externalities from their party being in power. I rely on a Regression Discontinuity (RD) to investigate the effect of party incumbency.[27] Table 2.2 and Graph 2.2 display the result of the regression. We see that party incumbency does not have a significant effect on the candidates' probability of winning in the open-seat races for the U.S. House of Representatives.[28]

[insert Table 2.2 and Graph 2.2 here]

Figure 2.1 illustrates the distribution of candidates' vote share over within open-seat races between 1990 and 2014. Part a) of the figure compares the histogram of the vote shares with the normal distributions and part b) shows the difference of the vote share cumulative distribution from the theoretical normal cumulative distribution. As both graphs demonstrate, the distribution of the vote shares is almost perfectly normal.[29]

Table 2.1 presents summary statistics of the candidates' vote share. The total number of observations is 1,148, which represents 574 races in total. The mean of candidates' vote share is 48.53 and the standard deviation is 13.55. The skewness of the distribution is -0.08 that is very close to that of a normal distribution, and the kurtosis is equal to 3.73, which

[24] In total 19 races did not satisfy this condition. Including these races in the regressions will not change the significance of the results and the change in the magnitude of the point estimates is negligible.

[25] The results of the regressions are robust to including these races. Seven races are excluded on this basis. Based on the paper by Snyder (1990), the races in which third parties won more that 10 percent of the votes have a different dynamic and are not comparable with other races.

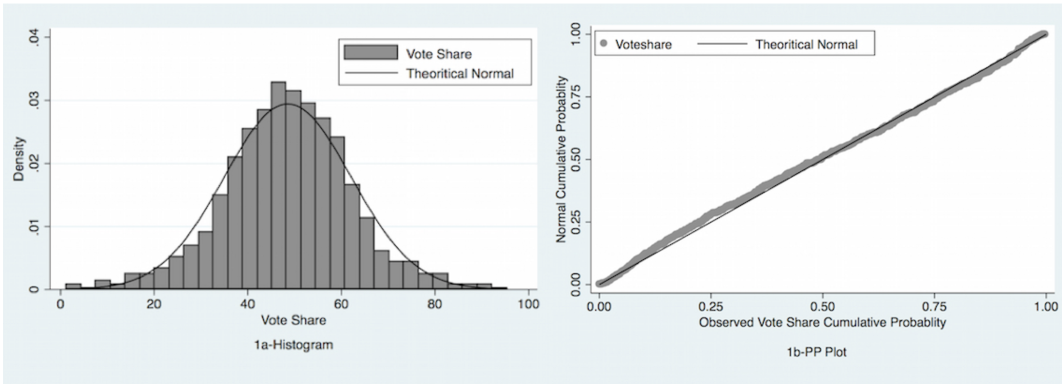
[26] For a complete survey see Hainmueller and Kern (2008).

[27] RD design is a technique to estimate the treatment effect. In this setting assignment of the treatment effect discontinuously changes as a function of one or more observable variables (Lee and Lemieux, 2010).

[28] The full description and results of the regression is shown in the appendix.

[29] It is worthwhile to mention that common normality tests are not informative in this case as the interval of vote share is limited (between 0 and 100). In this case, the third central moment (Skewness) and the fourth moment (Kurtosis) can be used to compare the probability distribution with normal distribution.(Greene, 2003, p. 1018).

Figure 2.1: The distribution of the vote share of the Democrat and Republican candidates in open-seated races from 1990 to 2014



Source: Federal Election Commission

is higher than the normal kurtosis. This implies that the tails are thicker than a normal distribution,[30] because the vote share distribution is limited between zero and 100.[31]

Table 2.1: Summary statistics of the candidates' vote share in open-seated races between1990 and 2014

Observation	Mean	Median	Standard Dev.	Skewness	Kurtosis
1148	48.63	48.8	13.42	-0.04	3.69

Source: Federal Election Commission

I consider races in which one of the candidates received more than 75.55 percent (or less than 21.71) of votes (that is the average of vote shares plus twice of the standard deviation) as outliers. The paper by Snyder (1990) argues that only in close races do candidates maximize contributions by promising the maximum number of favors that they can, as they are equally likely to win the race.[32] There is no unique definition of a close-race, but based on the normal distribution of the candidates' vote share, I define a close race as one in which the candidates' vote share lie within two standard deviations of the mean,[33] and investigate

[30]Heavy-tailed.

[31]According to Greene (2003) the Skewness of a normal distribution is equal to zero, and the kurtosis is equal to 3.

[32]As I show latter, the results of regression are robust to the definition of a close race.

[33]1,084 observations (94.4 percent) lie in the given interval.

if the closeness of the race affect the results of the regressions.

Table 2.4 presents summary statistics for the main variables. The average concentration ratio is 53.12% and the standard deviation is about 24%. As Table 2.4 and Figure 2.3 show, most of the industry contributions to candidates have been made by firms with a concentration ratio between 10 and 20% (Category 2). The three main participants in Category 2 are labour unions, law firms and attorneys, and physicians. Category 9 and 10 contribute the least amounts. Major contributors in these two categories are: pharmaceutical manufacturing and wholesale, investment banking, automotive manufacturers, telephone and cellphone providers, express delivery services, and air transportation.

[insert Table 2.4 and Figure 2.3 here]

2.5 Methodology

I use a linear probability model to investigate the effect of individual industries' contribution on the candidates' probability of winning. I estimate the following equation via Ordinary Least Squares (OLS):

$$Win_{it} = \beta_0 + Category_{it}\beta_1 + \beta_2 Amount_{it} + \beta_3 Popularity_{it} + \psi_t + \phi_r + \epsilon_{it}, \tag{2.1}$$

The dependent variable Win_{it} is a binary variable that takes the value of one if if candidate i wins, and zero otherwise.

$Category_{it}$ is a vector of interest that includes 10 dummies. Each element of this vector, $Category_{jit}$, is a dummy which is one if Candidate i have received money from the firms operating in Category j of industries, otherwise zero. j is an indicator of the degree of concentration of the industries. I categorized the industries based on their concentration ratio to 10 groups.[34] Index j is incremented by one for every 10 percentage point increase in the measure of concentration. For example, Category 1 represents the industries with a concentration ratio between 0 and 10 (which includes contributions made by firms in a very competitive market) and the last group is associated with the concentration ratio between 90 and 100 (firms that operate in a concentrated market).

It is well-documented that money has a positive impact on the probability of winning.[35] In order to control for the influence of total contributions received by a candidate, I include $Amount_{it}$ which is the total money that a candidate i received from all industries.

Although the advantage of candidate incumbency has been eliminated by excluding the races with incumbents, in some states candidates representing a specific party have a higher probability of winning. I use the variable $Popularity_{it}$ to capture the effect of the popularity of a party in a state. This variable, which was originally introduced by Snyder (1990), captures the average of the difference between the state and nationwide vote share of the past two presidential elections.[36]

ψ_t is a vector of time fixed effects to capture common trends across time. The vector ϕ_r includes race fixed effects to capture the unobservable factors within a race that might have affected the election results. Finally, ϵ_{it} is the error term. In the preferred model, error terms are corrected for heteroscedasticity using the HC3 method suggested by Davidson and MacKinnon (1995). For the purpose of robustness I also examined different ways of clustering the error terms.

[34] As it was mentioned before I use "the percentage share of the 50 largest firms of total sales, receipts, or revenue" as a measure of concentration, so the measure lies between 0 and 100.

[35] The book by Johnston (2014) presents a comprehensive review of this matter.

[36] As an example to clarify the way of calculating party strength. Consider a Republican candidate running in California in 2010, the popularity variable is calculated as follows:

$$Popularity_{it} = \sum_{T=2004}^{2008} \frac{VSR_{it} - NVSR_{it}}{2}, \tag{2.2}$$

where VSR_{it} is the vote share of the Republican presidential candidate at time t in California and $NVSP_{it}$ is the national vote share of the Republican presidential candidate.

2.6 Results

2.6.1 Main results

The main results of the paper are summarized in Table 2.5. Column 1 is the preferred specification, which considers all races and incorporates race and cycle controls. The results show that the total money that a candidate receives as campaign contributions and the party popularity both have a significant and positive effect on the candidate's probability of winning. The results imply that, on average, a 1 percent increase in the measurement of the party strength increases the probability of a candidate winning by 0.87%, and similarly, a one million dollars increase in the total amount of investor-contributions received by a candidate from profit seeking-firms increases the probability of winning by 82.03%. To put this in perspective, based on the summary statistics, a candidates receives on average 0.44 million dollars of investment-contributions. Those findings are similar to those of (Snyder, 1993, 1990).

More importantly, the results suggest that although the total amount of money matters, the sources of the contributions are important too. The positive and significant point estimates of Category 9 and Category 10 shows that within a specific race, the candidate who receives money from more concentrated industries has a higher probability of winning.[37] All else equal, raising money from industries in Category 10 (9) will increase the probability of the winning by 39.5% (22.1%). These results confirm the prediction of the collective action theory by Olson (2009) indicating that free-rider problems increase as the number of competitors grows. Moreover, as Figure 2.5 illustrates, the average firm size increases with the level of concentration of industries.[38] Grossman and Helpman (2001) argue that election outcomes depend on the turnout rate among different special interest groups. The results suggest that not only do concentrated industries suffer less from free-rider problems; they

[37] As it has been mentioned before the concentration ratio is "the percentage share of the 50 largest firms in an industry of total sales, receipts, or revenue of the industry".

[38] The average firm size is calculated by dividing the total number of employees by the total number of establishment in that industry.

also take advantage of economies of scale as they have relatively more employees. Information is a public good; gathering information is costly while the cost of transferring the knowledge to another person is negligible by comparison. Therefore, firms with a higher number of employees have more incentive to conduct research and inform their members about the preferred candidate, as it is more likely that the positive spillover of the research outweighs its costs. This result is consistent with the empirical result of the paper by Salamon and Siegfried (1977), which indicates that larger firms are more efficient in using their political power.

[insert Table 2.6 here]

2.6.2 Assessing the relative contribution of the components of the contributions on the election outcomes

I use standardized regression coefficients and perform a Shorrocks-Shapley decomposition of the R-squared in order to assess the relative contribution of each factor to the probability of a candidate winning.

Table 2.9 presents the standardized estimated coefficients, and I include close races for comparison. The standardized estimation of the preferred model suggest that a one standard deviation increase in the total amount of contributions received by a candidate will improve the chance of being elected by 27.97 percent,[39] while a standard deviation increase in the popularity of the party is associated with a 0.82 percent[40] increase in the probability of the winning of a candidate. Similarly, receiving money from industries in Category 9(10) will increase the chance of winning the office by 17.49(31.25) Percent.

[insert Table 2.9 here]

Table 2.10 shows the Shapley decomposition results for the total amount and components of the contributions. The results shows that the total amount of contributions explains 25.42

[39] (0.41*68.23).
[40] (0.069*0.122).

percent of the variability in the candidates' probability of winning, while the popularity of the party only explains 7.2 percent of the variation. The results suggest that the sources of the contributions explain 37.57 percent of the variation, and most of the variation is explained by receiving money from more concentrated firms. Receiving money from firms in Category 9 and 10 explains 29.54% for the total variation. The rest of the variation is explained by cycle and race dummies.

[insert Table 2.10 here]

Both standardized coefficients and Shapley decomposition results suggest that, controlling for the total amount and party strength, receiving money from more concentrated industries improves the candidate's probability of winning.

2.6.3 Robustness

Tables 2.5, 2.6, 2.7, and Figure 2.4 report the results of a number of robustness tests.

Columns 2, 3, and 4 of Table 2.5 illustrate the results using the open-seat races with different sets of dummies. Column 2 shows the results of the estimation using dummies for races; column 3 shows the results of the estimation of the same model where only cycle dummies are used; and in the estimates presented in column 4 neither race nor cycle dummies are used. As presented, the main results of the model stay intact in all four specifications, suggesting the estimates are robust.

Table 2.6 reports the results of the same model using only close open-seat races.[41] The specification of the estimations in columns 1, 2, 3, and 4 are similar to Table 2.5.[42] As presented, the estimated coefficients retain sign and significance and values only change slightly.[43] Figure 2.4 plots the results of the preferred specification of the model restricting

[41] In which the vote share of candidates are between 21.71 and 75.55.

[42] Column 1 shows the results of the model including both race and cycle dummies, column 2 shows the results of the estimation using only race dummies; column 3 displays the results of the estimation of the same model where only cycle dummies are used; and in column 4 neither race nor cycle dummies are used.

[43] In other words, using different specification, the point estimate of Category 9 varies within [0.23 0.16] and Category 10 changes less than 0.04 units.

the intervals of candidate's vote share. As depicted the change in the magnitude of the point estimates of Category 9 and 10 is negligible. Moreover, the point estimates of Category 10 remain statistically significant. The results of the estimations suggest that the results are not being driven by a specific definition of a close race. This is particularly important as Snyder (1993) argues that only in close races do candidates maximize the investor contributions, so the results suggest that regardless of how close the race is, controlling for the total money and the popularity of the candidate, the candidate who receives contributions from concentrated industries has a higher probability of winning.

[insert Table 2.6 and Figure 2.4 here]

Column 2 of Table 2.7 displays the results of the preferred model using total investment and non-investment contributions received by candidates. Even after taking into account the total contributions, receiving money from firms in Category 9 and 10 has a positive and significant effect on the odds of candidate's winning an election. Column 3 reports the results of the model using the exact measure of the concentration ratio (instead of the average). This excludes the races before 1997 (215 races) as the economic census started reporting the detailed data based on the NAICS codes only after 1997. As depicted the sign and significance of the Category 9 and 10 remain the same,[44] indicating that the results of the model is not driven by using the average concentration ratio.

There is no state-wide measure of industry concentration ratio available in the Economic Census. Based on the U.S. census Bureau, the states are categorized into 4 different regions; Northeast, Midwest, South, West.[45] To regionalize the concentration ratio, I multiple the national-wide concentration ratio by the ratio of region-wide per capita number of establishments to the national-wide per capita number of establishments, and categorize the contributions based on the regionalized concentration ratio.[46] The results of this exercise are

[44]Category 9 is significant at the 90 percent confidence interval.

[45]Information available at `https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf`.

[46]The races before 1997 are excluded as the detailed industry information is not available based on the NAICS codes.

reported in column 4 of Table 2.7. When using the regionalized concentration ratio the point estimate of Category 10 retains sign and significance and the value changes modestly compared to the column 3, suggesting that the results are not sensitive to using the nation-wide measure of the concentration ratio.

[insert Table 2.7 here]

2.6.4 Alternative standard errors

Table 2.8 reports different ways of clustering the error terms. In the preferred specification, I use the method suggested by Davidson and MacKinnon (1995), known as HC3, to correct for possible heteroscedasticity in the model.[47] However, I present the results of using alternative methods in Table 2.8. Column 2 represents estimates using heteroscedasticity-consistent Eicker-White standard errors, while in column 3 and 4 error terms are clustered at race and cycle level, respectively. In all the races the coefficients of the estimations are unchanged. Although the level of significance of some variables changes, the coefficients remain significant.

[insert Table 2.8 here]

2.7 Conclusion

Special interest groups use their political power to either gain some personal favors (influence channel) or to promote their favorite candidate (electoral channel). Delivering promised favors in return for contributions made by special interest groups are contingent on winning an election.[48] Therefore, SIGs try to affect election outcomes by financing the campaign

[47]Long and Ervin (2000) examines different possible ways of correcting for heteroscedasticity and suggests that HC3-based method is the most reliable way when estimating an OLS regression.

[48]Following the paper Baron (1994) and Snyder (1993), and considering the limitation imposed on the individual contributions by federal campaign finance laws, I assume that individual contributors are not looking for a change in the major public policies which might affect the decision of other informed voters.

expenses of their favorite candidates and by persuading their members to turn out to vote for the candidates which their interests are aligned. To evaluate the impact of market structure on the political power of the SIG, this paper utilizes data on the election of the United States House of Representatives between 1990 and 2014.

Even when controlling for total contributions received by a candidate and the popularity of his or her party, the results indicate that receiving contributions from more concentrated industries further improves the odds of winning. Within a race, receiving money from firms in Category 9(10)[49] increases the probability of winning by 17.49(31.25) percent, respectively. This is consistent with the prediction of the collective action theory by Olson (2009). The results also confirm that firms in more concentrated industries have a relatively higher number of employees per establishment, which is consistent with the prediction of the model by Grossman and Helpman (2001) indicating that larger SIGs have more political influence (considering both influence and electoral channel), because they benefit from economies of scale. The Shorrocks-Shapley decomposition results show that the components of individual-contributions are the most important factors in explaining the variation in candidates' probability of the winning. The results of the paper shows that it makes no difference whether the race is close or not, and the results prove robust to a variety of alternative specifications.

[49]Representing the industries in which the 50 largest firms receive 80 to 90(90 to 100) percent of the total industry sales, respectively.

Appendix

RD design

This paper uses a Regression Discontinuity (RD) design to investigate the effect of party incumbency on the probability of a candidate winning. The RD design was first developed by Thistlethwaite and Campbell (1960) to investigate the effect of merit awards on the later performance of the students. Since then, the design has been used widely in economics.[50] The RD design is applicable when the treatment effect is assigned based on a continuous and observable variable of interest which jumps over a certain threshold. For example in the original model by Thistlethwaite and Campbell (1960), the award was given to students who scored higher than a cut-off on a test. The model assumes that the assignment around the threshold is random, (which means there is no significant difference between the ability of the students) and investigates the effect of treatment by only comparing individuals around the cut-off.

In this paper, I use the difference between the vote share of a Democrat candidate and a Republican candidate at time t-1 as the score variable (Democrat party wins the election if this variable is positive). I run the regression for the Democrat candidates, but running the same regression for Republican party does not change the results. For a number of races the winning candidate received less than 50 percent of votes. To be consistent in assigning the treatment effect, I calculate the difference in the vote share of the Democrat and Republican candidates. The Democrat candidate wins the election if the difference is equal or greater than zero. The idea is when the previous candidate gets more than 50 percent of the vote she wins the race. I estimate the following model:

$$DifVoteShare_{dt} = \alpha_0 + \beta DifVoteShare_{d,t-1} + \gamma DemWin_{d,t-1} + \eta_{dt},$$

where $DifVoteShare_{dt}$ is the difference between the vote share of the "Democrate"

[50] For a full description see Lee and Lemieux (2010).

candidate and the "Republican" candidate at time t, $DifVoteShare_{p,t-1}$ is the difference between the vote share of the "Democrate" candidate and the "Republican" candidate at time $t-1$, η_{pt} is the error term, and $DemWin_{p,t-1}$ is defined as follow:

$$DemWin_{d,t-1}=\begin{cases} 1 & \text{if} \quad DifVoteShare_{D,t-1} \geq 0. \\ 0 & \text{if} \quad DifVoteShare_{D,t-1} < 0. \end{cases}$$

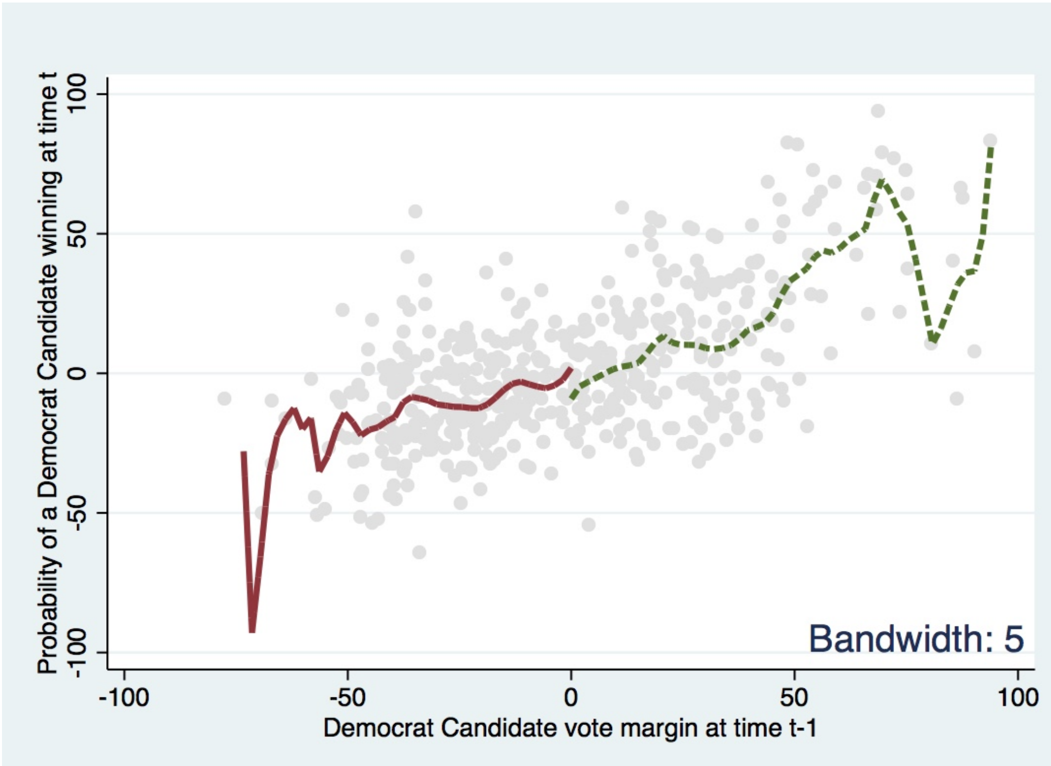
Table 2.2 and Graph 2.2 display the result of the regression. Table 2.2 represents both linear and polynomial estimation with three different bandwidths. As can be seen incumbency of the party does not have a significant effect on the candidate's probability of winning. Graph 2.2 confirms the results of the regression.

Table 2.2: The effect of party winning at t-1 on the party vote margin at time t (using regression discontinuity)

	Local Linear			Polynomial		
Bandwidth	5	10	15	5	10	15
$DemWin_t$	-11.08	-6.42	-4.085	-2.73	-0.966	-7.135
	(8.79)	(5.93)	(5.38)	(4.30)	(4.33)	(5.24)
	[0.207]	[0.279]	[0.448]	[0.526]	[0.823]	[0.173]

Notes: The dependent variable is the vote share of the candidate from the Democrat party in open-race elections of the U.S. House of Representative between 1990-2014. The explanatory variable is vote share of the candidate from the Democrat party in the previous election. The table shows both linear and polynomial RD setup with three different bandwidths (5, 10, and 15). Standard errors in parentheses. P-values from the goodness-of-fit test in square brackets. * significant at 5% ** significant at 1%, *** significant at 0.1%

Figure 2.2: Regression discontinuity design of the open-race election of the U.S. House of Representative between 1990-2014



Note: Simple Linear RD Setup. The dependent variable is the vote share of the candidate from the Democrat party in open-race elections of the U.S. House of Representative between 1990-2014. The explanatory variable is vote share of the candidate from the Democrat party in the previous election.

Table 2.3: Contribution limits (2013-2014)

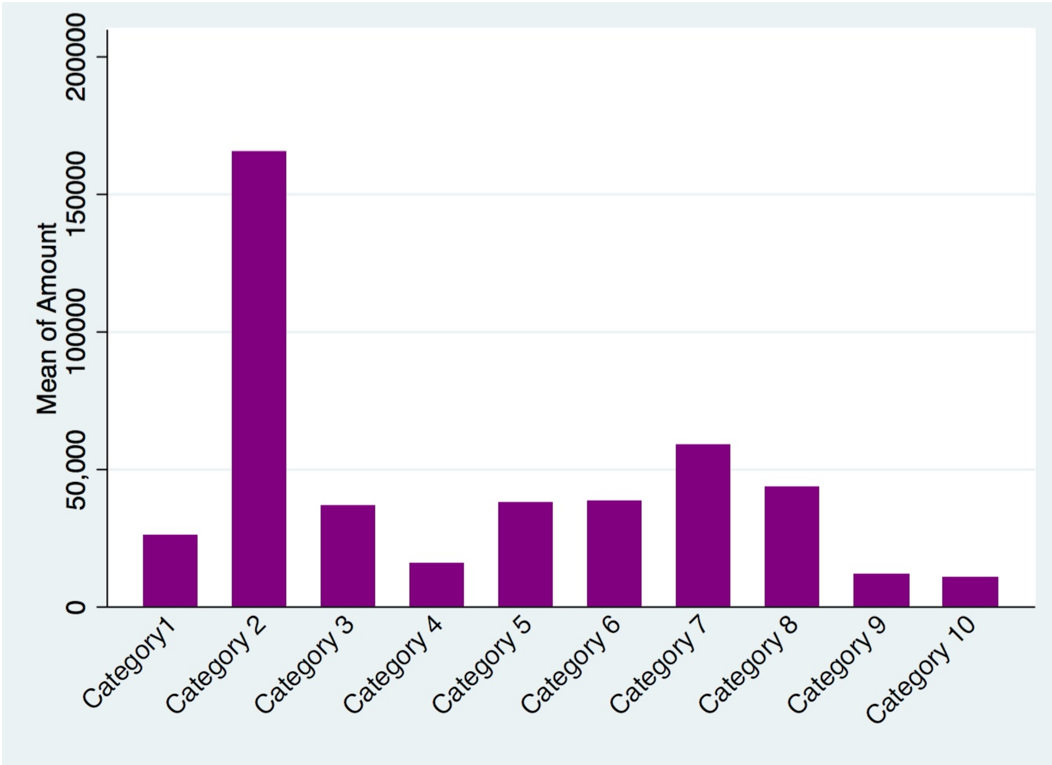
	To each candidate or candidate committee per election year	**To national party committee per calendar year**	**To state, district, and local party committee calendar year**	**To any other political committee per calendar year**	**Special limit**
Individuals	2,600	32,400*	10,000 (combined)	5,000	No limit
National Party Committee	5,000	No limit	No limit	5,000	45,000* to Senate candidate per campaign
State, District, and Local Party Committee	5,000 (combined)	No limit	No limit	5,000 (combined)	No limit
PAC (multicandidate)	5,000	15,000	5,000 (combined)	5,000	No limit
PAC (not multicandidate)	2,600*	32,400*	10,000 (combined)	5,000	No limit

Source: FECA 2014
*These contribution limits have been incremented according to the inflation in odd-numbered years since 2005.

Table 2.4: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
Part a				
Total Amount(millions of dollars)	0.4434	0.4161	0	2.8230
Total Amount Category 1	25,811	30,249.27	0	249,705
Total Amount Category 2	165,436.33	169,121.69	0	1,231,237
Total Amount Category 3	36,782.17	42,496.31	0	302,500
Total Amount Category 4	15,636.62	20,347.73	0	193,250
Total Amount Category 5	37,460.26	46,682.8	0	356,750
Total Amount Category 6	38,328.95	40,694.28	0	320,520
Total Amount Category 7	58,821.06	73,972.40	0	455,024
Total Amount Category 8	43,063.12	52,469.08	0	511,027
Total Amount Category 9	11,431.81	18,547.79	0	228,650
Total Amount Category 10	10,647.99	13,721.23	0	172,351
Amount Dummy 1	0.622	0.485	0	1
Amount Dummy 2	0.678	0.467	0	1
Amount Dummy 3	0.638	0.481	0	1
Amount Dummy 4	0.586	0.493	0	1
Amount Dummy 5	0.632	0.482	0	1
Amount Dummy 6	0.634	0.482	0	1
Amount Dummy 7	0.642	0.48	0	1
Amount Dummy 8	0.627	0.484	0	1
Amount Dummy 9	0.55	0.498	0	1
Amount Dummy 10	0.553	0.497	0	1
Party Strength	0.0003	0.0698	-0.422	0.422
Candidate's Vote Share	48.63	13.43	1.2	95.40
Number of Observations	1148			
Number of Races	574			
Part b				
Concentration Ratio	53.12	23.703	1.268	100
Number of Establishments	60,782.312	112,397.933	30	888,229
Number of Employees	930,888.099	1,681,330.624	0	16,945,834
Firm Size	31.517	37.169	0	296.719

Figure 2.3: Total amount received by candidates (1990-2014)



Note: This figure plots the total amount that each category of firms contributed to candidates of open-seat races for the U.S. House Representatives between 1990 and 2014. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration.

Table 2.5: The effect of the sources of campaign contributions on the odds of a candidate winning an open-seat race between 1990 and 2014

	(1) Preferred	(2) RaceDummy	(3) CycleDummy	(4) NoDummy
Party Strength	0.8719***	0.8719***	1.3813***	1.4082***
	[0.2973]	[0.2973]	[0.1738]	[0.1745]
Total Amount	0.8203***	0.8203***	0.3067***	0.2198***
(millions of dollars)	[0.1098]	[0.1098]	[0.0394]	[0.0341]
Category 1	-0.0128	-0.0128	0.0180	0.0177
	[0.1243]	[0.1243]	[0.0508]	[0.0481]
Category 2	-0.1235	-0.1235	-0.0439	-0.0326
	[0.1227]	[0.1227]	[0.0441]	[0.0409]
Category 3	-0.0197	-0.0197	-0.0066	-0.0380
	[0.1297]	[0.1297]	[0.0487]	[0.0470]
Category 4	0.0832	0.0832	0.0719	0.0517
	[0.1242]	[0.1242]	[0.0488]	[0.0487]
Category 5	-0.0188	-0.0188	-0.0174	-0.0401
	[0.1181]	[0.1181]	[0.0452]	[0.0433]
Category 6	0.0596	0.0596	-0.0138	-0.0168
	[0.1093]	[0.1093]	[0.0403]	[0.0380]
Category 7	0.0922	0.0922	-0.0076	-0.0043
	[0.0973]	[0.0973]	[0.0426]	[0.0403]
Category 8	-0.1363	-0.1363	-0.0680	-0.0279
	[0.1093]	[0.1093]	[0.0447]	[0.0418]
Category 9	0.2214**	0.2214**	0.1560***	0.1891***
	[0.0992]	[0.0992]	[0.0419]	[0.0405]
Category 10	0.3954***	0.3954***	0.3251***	0.3488***
	[0.0825]	[0.0825]	[0.0349]	[0.0341]
Race Dummy	Y	Y	N	N
Cycle Dummy	Y	N	Y	N
Observations	1148	1148	1148	1148

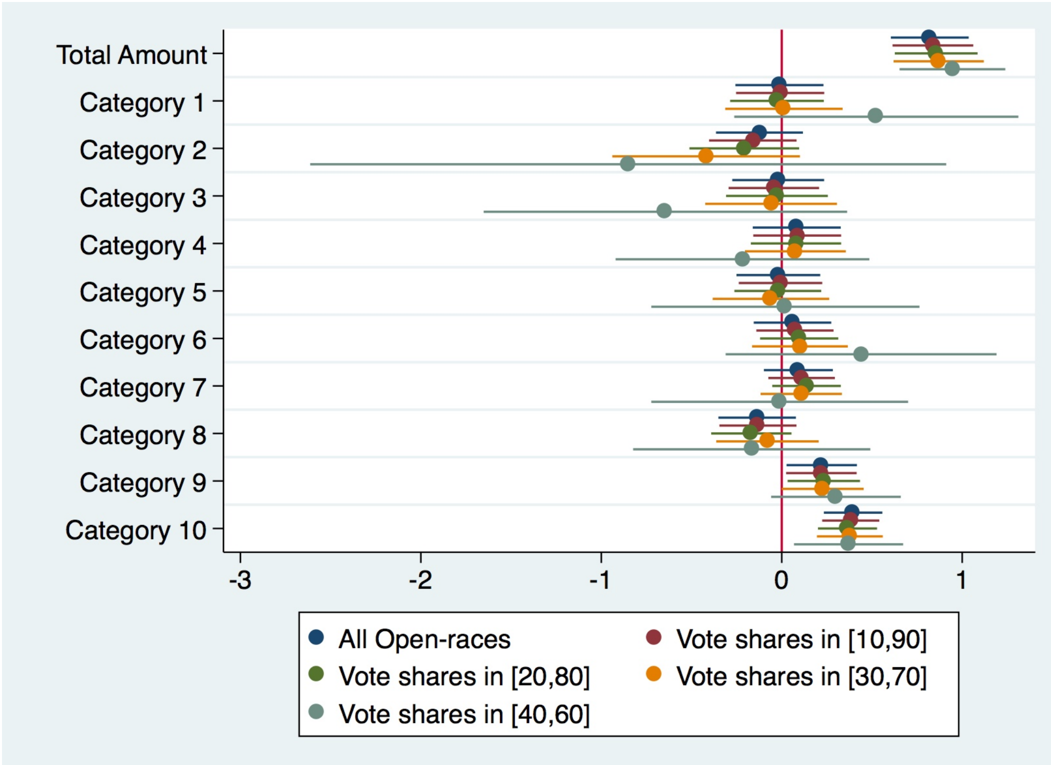
Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Robust standard errors are in parentheses. The estimation includes the open-seats races for the U.S. House Representatives. Column 1 displays the results of regression using both cycle and race dummies, Column 2 shows the results of the estimation using dummies for races; Column 3 shows the results of the estimation of the same model where only cycle dummies are used; and in the estimation of the results presented in Column 4 neither race nor cycle dummies are used. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 2.6: Robustness: Close open-seat races

	(1) Close Races	(2) RaceDummy	(3) CycleDummy	(4) NoDummy
Party Strength	0.9310***	0.9310***	1.3522***	1.3783***
	[0.3123]	[0.3123]	[0.1812]	[0.1820]
Total Amount	0.8697***	0.8697***	0.3293***	0.2320***
(millions of dollars)	[0.1195]	[0.1195]	[0.0407]	[0.0349]
Category 1	-0.0227	-0.0227	0.0209	0.0210
	[0.1362]	[0.1362]	[0.0525]	[0.0493]
Category 2	-0.2225	-0.2225	-0.0226	-0.0114
	[0.1741]	[0.1741]	[0.0540]	[0.0541]
Category 3	-0.0300	-0.0300	-0.0130	-0.0498
	[0.1482]	[0.1482]	[0.0524]	[0.0511]
Category 4	0.0777	0.0777	0.0630	0.0378
	[0.1290]	[0.1290]	[0.0499]	[0.0497]
Category 5	0.0028	0.0028	-0.0184	-0.0431
	[0.1280]	[0.1280]	[0.0479]	[0.0461]
Category 6	0.0785	0.0785	-0.0039	-0.0126
	[0.1147]	[0.1147]	[0.0421]	[0.0396]
Category 7	0.1493	0.1493	0.0001	0.0019
	[0.0977]	[0.0977]	[0.0438]	[0.0414]
Category 8	-0.1521	-0.1521	-0.0905*	-0.0333
	[0.1155]	[0.1155]	[0.0471]	[0.0441]
Category 9	0.2319**	0.2319**	0.1548***	0.1899***
	[0.1035]	[0.1035]	[0.0420]	[0.0408]
Category 10	0.3515***	0.3515***	0.3143***	0.3426***
	[0.0852]	[0.0852]	[0.0346]	[0.0337]
Race Dummy	Y	Y	N	N
Cycle Dummy	Y	N	Y	N
Observations	1084	1084	1084	1084

Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Robust standard errors are in parentheses. The estimation includes the close open-seats races for the U.S. House Representatives. Column 1 displays the results of regression using both cycle and race dummies, Column 2 shows the results of the estimation using dummies for races; Column 3 shows the results of the estimation of the same model where only cycle dummies are used; and in the estimation of the results presented in Column 4 neither race nor cycle dummies are used. * significant at 10% ** significant at 5%, *** significant at 1%.

Figure 2.4: Closeness of the races



Note: This figure plots the coefficients for the total amount received by a candidate and the source of money based on the concentration ratio. The regressions have been run for different intervals of candidates' vote share. Lines show the 95 percent confidence interval based on robust standard errors. The dependent variable is a dummy for winning an election. Total amount is in millions of dollars. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. All the regressions include controls for race, cycle and popularity of the party.

Table 2.7: Robustness

	(1) Preferred	(2) All Contributions	(3) Yearly Concentration Ratio	(4) Region Based Concentration Ratio
Party Strength	0.8719***	0.8424***	1.0050***	1.0776***
	[0.2973]	[0.3224]	[0.3143]	[0.3156]
Total Amount	0.8203***		0.8510***	0.9196***
(millions of dollars)	[0.1098]		[0.1183]	[0.1143]
Total Amount (including		0.3785***		
non-investment)(millions of dollars)		[0.0954]		
Category 1	-0.0128	-0.0129	0.0051	-0.0706
	[0.1243]	[0.1251]	[0.1128]	[0.1117]
Category 2	-0.1235	-0.1809	-0.1766	-0.2116*
	[0.1227]	[0.1478]	[0.1266]	[0.1261]
Category 3	-0.0197	-0.0450	0.0243	-0.1001
	[0.1297]	[0.1253]	[0.1269]	[0.1443]
Category 4	0.0832	0.0913	0.0259	0.0761
	[0.1242]	[0.1237]	[0.1099]	[0.1068]
Category 5	-0.0188	-0.0567	0.0010	0.0073
	[0.1181]	[0.1194]	[0.1148]	[0.1144]
Category 6	0.0596	0.0985	0.0883	0.0694
	[0.1093]	[0.1368]	[0.0989]	[0.1045]
Category 7	0.0922	0.1427	0.0324	0.1275
	[0.0973]	[0.1290]	[0.1128]	[0.1078]
Category 8	-0.1363	-0.1503	-0.0721	0.0807
	[0.1093]	[0.1123]	[0.1073]	[0.1149]
Category 9	0.2214**	0.2733**	0.2025**	0.1692
	[0.0992]	[0.0975]	[0.1176]	[0.1253]
Category 10	0.3954***	0.4638***	0.2781***	0.2166**
	[0.0825]	[0.0874]	[0.1044]	[0.1089]
Race Dummy	Y	Y	Y	Y
Cycle Dummy	Y	Y	Y	Y
Observations	1148	1148	718	718

Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Clustered standard errors are in parentheses. The estimation includes the open-seats races for the U.S. House Representatives. Both cycle and race dummies are used in all the estimations. Column 1 represents the preferred specifications. Column 2 shows the results of the estimation using the total investment and non-investment contributions. Column 3 displays the results of the preferred model using census-year concentration ratios. In Column 4 census-year concentration ratios have been regionalized based on the number of establishments in each census region. * significant at 5% ** significant at 1%, *** significant at 0.1%.

Table 2.8: Alternative standard errors

	(1) Robust HC3	(2) Robust White	(3) Clustered at Race level	(4) Clustered at Cycle Level
Party Strength	0.8719***	0.8719***	0.8719***	0.8719**
	[0.2973]	[0.2082]	[0.2946]	[0.3443]
Total Amount	0.8203***	0.8203***	0.8203***	0.8203***
(millions of dollars)	[0.1098]	[0.0757]	[0.1072]	[0.1430]
Category 1	-0.0128	-0.0128	-0.0128	-0.0128
	[0.1243]	[0.0841]	[0.1190]	[0.1653]
Category 2	-0.1235	-0.1235	-0.1235	-0.1235
	[0.1227]	[0.0829]	[0.1173]	[0.1181]
Category 3	-0.0197	-0.0197	-0.0197	-0.0197
	[0.1297]	[0.0880]	[0.1246]	[0.0966]
Category 4	0.0832	0.0832	0.0832	0.0832
	[0.1242]	[0.0857]	[0.1212]	[0.0982]
Category 5	-0.0188	-0.0188	-0.0188	-0.0188
	[0.1181]	[0.0803]	[0.1135]	[0.1219]
Category 6	0.0596	0.0596	0.0596	0.0596
	[0.1093]	[0.0739]	[0.1045]	[0.1158]
Category 7	0.0922	0.0922	0.0922	0.0922
	[0.0973]	[0.0655]	[0.0926]	[0.1208]
Category 8	-0.1363	-0.1363*	-0.1363	-0.1363
	[0.1093]	[0.0738]	[0.1044]	[0.1019]
Category 9	0.2214**	0.2214***	0.2214**	0.2214***
	[0.0992]	[0.0688]	[0.0973]	[0.0695]
Category 10	0.3954***	0.3954***	0.3954***	0.3954**
	[0.0825]	[0.0570]	[0.0807]	[0.1661]
Race Dummy	Y	Y	Y	Y
Cycle Dummy	Y	Y	Y	Y
Observations	1148	1148	1148	1148

Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Clustered standard errors are in parentheses. The estimation includes the open-seats races for the U.S. House Representatives. Both cycle and race dummies are used in all the estimations. In Column 1 error terms are corrected for heteroscedasticity using the method suggested by Davidson and MacKinnon (1995). Column 2 represents heteroscedasticity-consistent Eicker-White standard error, in Column 3 and 4 error terms at clustered at race and cycle level respectively.* significant at 10% ** significant at 5%, *** significant at 1%.

Table 2.9: The effect of the components of campaign contributions on the odds of a candidate winning an open-seat race (based on the market structure)

	(1)		(2)	
	Base(All Open-Seat Races)		Close Open-Seat Races	
	β Coefficient	Standardized β	β Coefficient	Standardized β
Party Strength	0.8719***	0.1217***	0.9310***	0.1294***
	[0.2973]		[0.3123]	
Total Amount	0.8203***	0.6823***	0.8697***	0.7263***
(millions of dollars)	[0.1098]		[0.1195]	
Category 1	-0.0128	-0.0079	-0.0227	-0.0130
	[0.1243]		[0.1362]	
Category 2	-0.1235	-0.0514	-0.2225	-0.0742
	[0.1227]		[0.1741]	
Category 3	-0.0197	-0.0113	-0.0300	-0.0157
	[0.1297]		[0.1482]	
Category 4	0.0832	0.0599	0.0777	0.0536
	[0.1242]		[0.1290]	
Category 5	-0.0188	-0.0111	0.0028	0.0015
	[0.1181]		[0.1280]	
Category 6	0.0596	0.034	0.0785	0.0411
	[0.1093]		[0.1147]	
Category 7	0.0922	0.051	0.1493	0.0753
	[0.0973]		[0.0977]	
Category 8	-0.1363	-0.0786	-0.1521	-0.0805
	[0.1093]		[0.1155]	
Category 9	0.2214**	0.1749**	0.2319**	0.1785**
	[0.0992]		[0.1035]	
Category 10	0.3954***	0.3128***	0.3515***	0.2694***
	[0.0825]		[0.0852]	
Race Dummy	Y	Y	Y	Y
Cycle Dummy	Y	Y	Y	Y
Observations	1148		1084	
R^2	0.52		0.505	

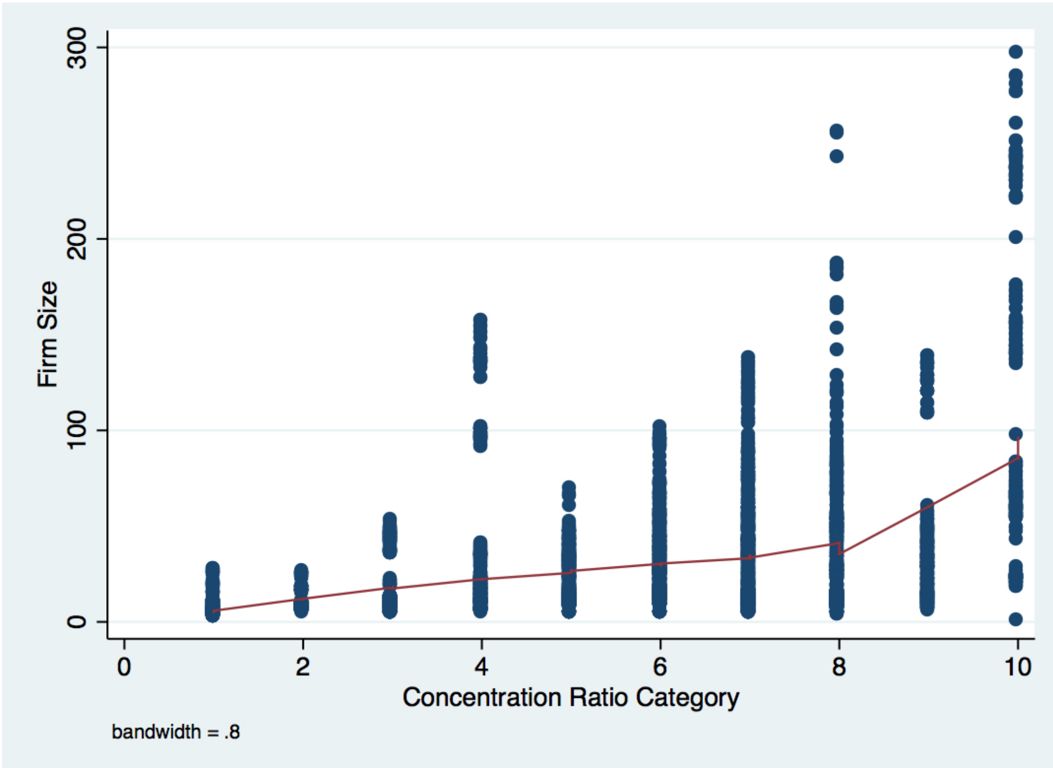
Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Robust standard errors are in parentheses. The first two columns include the open-seats races for the U.S. House Representatives, and the last two columns represent the close open-seats races for the U.S. House Representatives. Both cycle and race dummies are used in all the estimations. * significant at 5% ** significant at 1%, *** significant at 0.1%.

Table 2.10: Shorrocks-Shapley decomposition of the R-squared

	(1)		(2)	
	Base(All Open-Seat Races)		Close Open-Seat Races	
	Shapley Value	Shapley $\%R_2$	Shapley Value	Shapley $\%R_2$
Party Strength	0.0364	7.2	0.0387	7.42
Total Amount (millions of dollars)	0.1285	25.42	0.1238	23.79
Category 1 to 8	0.0406	8.03	0.0469	9.01
Category 9	0.061	12.09	0.0661	12.69
Category 10	0.0881	17.45	0.971	18.65
Cycle Dummies	0.0046	0.92	0.0043	0.82
Race Dummies	0.146	28.88	0.1437	27.61
Race Dummy	Y	Y	Y	Y
Cycle Dummy	Y	Y	Y	Y
Total	0.5053	100	0.5204	100

Notes: The dependent variable is a dummy for winning an election. Party strength is a state-wide measure of the popularity of the party. Category j is a binary indicator for whether candidate i received money from firms operating in category j. Index j is an indicator of the degree of concentration of the industries. Industries are categorized into 10 groups based on their concentration ratio. The indicator j is incremented by one with every 10 percent increase in the measure of concentration. Robust standard errors are in parentheses. The first two columns include the open-seats races for the U.S. House Representatives, and the last two columns represent the close open-seats races for the U.S. House Representatives. Both cycle and race dummies are used in all the estimations. * significant at 10% ** significant at 5%, *** significant at 1%.

Figure 2.5: The Relation between concentration ratio and the average size of firms operating in that industry



Note: This figure plots the correlation between average size of firm and the concentration ratio of the industry between 1997 and 2014. The average size of firm has been calculates by dividing the total number of employees by the total number of establishments in a specific industry. The concentration ratio is “the percentage share of the 50 largest firms of total sales, receipts, or revenue”.

Chapter 3

Ethnic Heterogeneity and STDs in Canada†

3.1 Introduction

Sexually Transmitted Diseases (STDs) continue to be an important issue of concern in Canada. The Public Health Agency of Canada reports that STDs account for more than 50 percent of all reported infectious diseases in Canada (PHAC, 2015). Despite all the efforts of Health Canada and the PHAC to promote the prevention of the STDs, rate of reported cases of chlamydia, gonorrhea, and infectious syphilis increased by 57.6, 38.9, and 101.0 percent, respectively, between 2003 and 2012.[1]

STDs impose a tremendous health burden on individuals. Untreated gonorrhea and chlamydia are the main cause of Pelvic Inflammatory Diseases (PID) (Wiesenfeld et al.,

†This research was supported by funds to the Canadian Research Data Centre Network (CRDCN) from the Social Science and Humanities Research Council (SSHRC), the Canadian Institute for Health Research (CIHR), the Canadian Foundation for Innovation (CFI), and Statistics Canada. Although the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada. In addition, the data includes information copied with permission from Canada Post Corporation.

[1]Public Health Agency of Canada (2015).

2012). PIDs can lead to serious and long-term medical complications such as to infertility, ectopic pregnancy, and chronic pelvic pain. Moreover, untreated STDs facilitate the sexual transmission of HIV infections by increasing the HIV viral load in genital fluids (Wasserheit, 1992). In addition to the potential long term effect of STDs on the health and well-being of individuals, they place a substantial economic strain on the health-care system of Canada. For instance, the expected annual cost of chlamydia, the most commonly reported notifiable infection in Canada (PHAC, 2012), is about 50 million dollars (Tuite et al., 2012), and the estimated cost of new HIV infections, which includes the present value of treatments, loss of productivity, and loss of quality of life, is about $1.3 million per person (Kingston-Riechers, 2011).

Identifying the determinant factors that contribute to the spread of STDs is crucial for preventing future infections and efficiently allocating scare health resources. In recent years, a wide strand of literature has focused on developing and applying mathematical models to understand, predict and prevent the spread of infectious diseases including STDs (Chowell and Hyman, 2016; Eames and Keeling, 2004; Zuma et al., 2003; Lin et al., 2013). However, modelling STDs epidemics remains challenging, as they differ from other infectious diseases in their unique method of transmission.[2] It is well documented that the structure of sexual network is a vital determinant of STD propagation (Keeling, 2005; Potterat et al., 2002; Pongou, 2009; Cahu et al., 2014; Pastor-Satorras et al., 2015).

In this paper we focus on ethnic heterogeneity as a determinant of STDs in Canada. One out of five residents of Canada is a foreign-born. According to the 2011 National Household Survey (NHS), Canada's proportion of foreign-born individuals, valued at 20.6 percent,[3] ranked first among the G8 countries.[4] Racial and ethnic ties create closed social networks with rigid in-group boundaries, hampering the intra-group communication and dissemination of information (McPherson et al., 2006; Kossinets and Watts, 2009). Slow information flow

[2]According to the World Health Organization (WHO, 2016), STDs mainly transmitted through sexual union with an individual who is carrying an STD.

[3]Approximately 6,775,800 of Canadian residences are foreign-born individuals.

[4]Including Canada, France, Germany, Italy, Japan, Russia, the United Kingdom, the United States.

among groups encourages individuals to ethnically diversify their sexual partners in order to lower the chance of getting caught cheating. Anonymity thus facilitates the spread of STDs by promoting sexual infidelity and broadening sexual networks among ethnic groups.

We empirically test the hypothesis of the paper using data collected from the 2013-2014 Canadian Community Health Survey (CCHS). This paper only includes the respondents of the survey aged between 15 and 49 who have had intercourse at least once in their life. We compute an index of ethnic heterogeneity, named HHI, at the forward sortation area level and find that ethnic diversity increases the number of sexual partners, and hence the chances of contracting a sexually transmitted disease. Our results are robust to a wide range of individual controls.

Estimating the effect of ethnic diversity on sexually transmitted diseases is challenging. The extent of ethnic diversity in a neighborhood might be driven by voluntary migration into the area. Individuals' characteristics such as family type, education, ability, and sexual desire, can highly impact the selection of a place to live. This in turn affects ethnic heterogeneity and at the same time the structure of sexual networks, and hence the chances of contracting a sexually transmitted disease. Therefore, the estimates of the paper can be biased since it is difficult to control for all the characteristics of the individuals. We tackle the potential endogeneity issue described above in three different ways.

First, following Altonji et al. (2005), we test the sensitivity of the model to the selection of unobservable variables, by using additional observable explanatory controls, including age, gender, marital status, sexual orientation, migration status, education, occupation, household income, and drinking and smoking habits. We find that the sign and magnitude of the HHI effect is remarkably stable when adding a rich set of controls. This insensitivity of the estimates suggests that the bias owing to unobservables is small, and it is therefore unlikely that the selection of unobservables explains away the entire effect of ethnic heterogeneity on the outcomes. Second, we instrument for the FSA-level value of ethnic heterogeneity by using the 2006, 2001, and 1996 values of HHI, average house prices, and average household income. The results indicates that the positive effect of HHI on the number of sexual partners and the prevalence of STDs is not due to recent voluntary migration into a neighborhood, nor it is related to the unobservable variables. Finally, we further confirm the validity of the casual HHI effect on outcomes by showing that ethnic heterogeneity does not affect the prevalence of non-transmitted diseases (e.g., asthma, arthritis, high blood pressure, heart disease, migraine, cancer, and diabetes).

The rest of the paper is laid out as follows. Section 2 surveys previous related literature. Section 3 explains the mechanism by which ethnic heterogeneity affects the formation of human sexual network and the prevalence of STDs. Next, we briefly discuss the measure of ethnic heterogeneity used in this paper. Section 5 introduces the data and the next section describes empirical approach of the paper. Section 7 discusses our results, and the last section concludes.

3.2 Literature review

To the best of our knowledge, this study is the first study to examine the effect of ethnic diversity on the spread of STDs in a multi-ethnic developed country. The theoretical framework of this paper in based on the paper by Pongou (2009) which proposes a theory that explains the role of ethnic diversity in formation of sexual networks, and investigates the effect of ethnic diversity on the spread of the sexually transmitted diseases in Africa.

Previous studies show a pattern of homophily in social and economic human networks (Currarini et al., 2009; McPherson et al., 2001; Jackson, 2010; Hsieh and Lee, 2016; Currarini et al., 2009; Golub and Jackson, 2012). By contrast, Pongou (2009) shows that this pattern does not work in socially prohibited relationships, such as sexual infidelity. Individuals diversify their sexual partners ethnically to hide behind the anonymity, and thus in addition to deriving utility from having multiple sexual partners, they benefit from the lower chances of getting caught cheating. This is aligned and adds to the previous studies by Pongou and

Serrano (2009, 2013).

This paper relates to a number of studies which show that understanding the structure of sexual networks is essential to explain the transmission of STDs. Liljeros et al. (2003) and Doherty et al. (2005) demonstrate that the dynamic of sexual partnerships and the topology of sexual networks are better predictors of the dissemination of STDs than individual behavior. In a related paper, Eames and Keeling (2004) show that even rare presence of mixed and concurrent sexual relationships in a network imposes a significant risk of infection on all members of that network, even though they are practicing serial monogamy. Furthermore, Pongou and Serrano (2009, 2013) show that the structure of a sexual network can explain the existing gap between male and female prevalence of HIV/AIDS. For instance, they show that in a patriarchal society, in which men hold the positions of power, women are more susceptible to HIV infection than men, and that women's vulnerability to infection with HIV decreases by progressing towards gender equality and women's empowerment.

The results of this study are consistent with a growing body of literature that explores the effect of migration as a social determinant of the spread of STDs. Mayer and Beyrer (2007) show that the recent increase in labour mobility in South Africa is associated with a higher rate of HIV infections. Particularly, they show that male migrants' risk of infection is 26 times larger in non-marital relationships with local females than in their relationships with their own wives. Hogben and Leichliter (2008) argue that in the process of settling and integrating into the social and economic realms of the United States, new migrants might engage in high-risk sexual behaviors (i.e., unprotected paid sex) and thus get infected. Later, they can act as a bridge among sexual networks to transmit the infection to members of other communities.

This paper also contributes to the strand of literature investigating the determinants of STDs in Canada. According to Wylie and Jolly (2001), the size of a sexual network varies between 2 and 82 individuals in Manitoba, and the chances of STDs transmission is significantly higher in groups with more participants. De et al. (2007), using information on

5,701 STD patients in Alberta between 1991 and 2003, show that ethnic identity and sexual orientation are the two main determinants of gonorrhea reinfection between Canadians. Sen and Luong (2008), employing cross-province data from 1991 to 1999, show that higher beer prices, as a proxy of costs of alcohol consumption, are associated with a reduction in the rate of chlamydia and gonorrhea infections in Canada. Bajaj et al. (2017) using a sample of 35,099 Canadian individuals, demonstrate that individuals' place of residence and behavioral factors (e.g., income, marital status, education, alcohol and smoking habits) are determinants of the STDs transmission in Canada.

3.3 Mechanism of influence

In this section, we provide the mechanism by which ethnic heterogeneity affects the formation of the sexual network structure in a multi-ethnical society and hence, is a determinant of the spread of sexually transmitted diseases.

Modelling sexually transmitted diseases is different from other contagious diseases, as they are uniquely transmitted through sexual union with a person who is carrying an STD. Although many recent studies have focused on the individuals' sexual behaviours as fundamental factors of the prevalence of STDs, the main determinant of STDs propagation at the population level is the structure of the sexual network. The implied network structure is a vital factor of determining the population patterns of exposure. Therefore, in the recent years, a growing body of literature has been devoted to identify the factors affecting the structure of the sexual networks. Although many factors affect the size, formation of linkages, and connectivity of a sexual network, this paper focuses on ethnic diversity as a determinant factor. In other words, the primary aim of this paper is to determine whether ethnic heterogeneity increases the number of sexual partner and the prevalence of STDs in a multi-cultural country.

Sexual networks refer to a group of individuals who are linked either directly or indirectly through sexual contact with each other (Flom et al., 2001). Many studies show that having multiple and concurrent sexual partner is the main building block of sexual networks that facilitates transmission of STDs (Eames and Keeling, 2004; Pongou and Serrano, 2013; Halperin and Epstein, 2004). According to Laumann (1994), Size, density, and connectivity of sexual networks are the main factors affecting the spread of STDs. Anderson (1999) shows that the presence of a small group of people who are not in a committed relationships and tend to change their partners frequently has significant implications for the prevalence of STDs in a society.

According to Morris and Kretzschmar (1995), as relationships can overlap or start without any delay, having multiple sexual partners permit a rapid spread of STDs through a network which leads to an increase in the number of individual who are infected at a later time. The paper by Pongou (2009) shows how ethnic heterogeneity can increase the optimal number of sexual partner and positively affect the transmission and persistence of a STD in a multi-ethical society. The paper assumes a society that is made up of several different ethnic groups, and regardless of ethnic background, individuals gain positive utility from having sexual partners. Finding a partner involves a positive cost, as it requires spending time and effort. The society does not appreciate polygamy and the probability of detecting infidelity is greater than zero. An unfaithful individual will be punished if caught cheating, the cost associate with infidelity has two major components: First, as monogamy is a norm in the society all couples, married or common law, consider sexual exclusively as a contractual condition of their union. By breaking the bargains, individuals are at risk of losing their family ties and relationships.[5] Second, given that monogamy is a social norm, having multiple sexual partners within a short period time is considered as a failure to commit and it is not appreciated by the members of the society. Therefore, the social reputation of the individuals with multiple sexual partners are impaired and they might get abandoned by the other members which will make it difficult for them to find a long-term partner in the

[5] Treas and Giesen (2000).

future. As a result, individuals have a strong preference for hiding the presence of multiple sexual partners. Ethic groups in a society create close social networks with pre-defined communication networks, hampering the intra-group communications and dissemination of information. This means information flows more easily within a group rather than intra-group (McPherson et al., 2006; Kossinets and Watts, 2009). The slow rate of information dissemination fosters anonymity across different ethnic groups which reduces the chance of being caught cheating, and therefore, leads to an increase in the individuals' optimal number of sexual partners. This optimizing behaviour increases the scope of a sexual network and facilitates the transmission of sexually infections.

3.3.1 Testable implications

As mentioned, ethnic heterogeneity is a determinant factor of the formation of human sexual networks. Ceteris paribus, introducing a new ethnic group will increase the individual's optimal number of partners and therefore affect the spread of sexually transmitted diseases. The network effect of the model has two testable implications:

1. Controlling for the individuals' characteristics, a person who lives in a more ethnically diverse neighborhood is more likely to have more sexual partners compared to one who lives in a homogenous neighborhood.

2. As discussed above, an increase in the number of sexual partners has a great implications for transmission of STDs in a society. Therefore it is expected that incrementing ethnic diversity increases the probability of catching an STD by positively affecting the optimal number of partners and broadening the scope of a sexual network.

3.4 Ethnic diversity

This study uses the ethnolinguistic fractionalization index (referred to as HHI)[6] as a measure of ethnic heterogeneity. This index measures the probability that two randomly chosen individuals from a population are from different ethnic groups. The HHI is the most frequently employed measure of ethnic fragmentation in the economic literature (Posner, 2004).

We calculate the index of ethnic heterogeneity, at the Forward Sortation Area (FSA) level.[7] Each FSA is composed of the first three characters of Canadian postal codes as follows: the first character is a letter that represent the province or territory (shown in figure 3.1), the second character is a number that shows if the area is urban or rural, and the last character is a more specific identifier of an urban or rural area. We use the following formula to calculate the ethnic diversity within each FSA:

$$HHI_f = 1 - \sum_e s_{ef}^2, \tag{3.1}$$

where s_{ef} denotes the share of ethnic group e in FSA f. Figure 3.2 shows the map of ethnic heterogeneity across Canada, and Figure 3.3 shows the distribution of the ethnic heterogeneity in the sample.

3.5 Data and empirical strategy

3.5.1 Data

The empirical analysis of this study uses data from the master files of the Canadian Community Health Survey (CCHS) 2013-2014, National Household Survey (NHS) 2011, and publicly available Censuses of Population 2006, 2001, and 1996: profile series. All dependent and explanatory variables are driven from the master file of the 2013-2014 CCHS except the main variable of interest, ethnic heterogeneity, which is calculated using the master file of the 2011

[6] It is also known as the ELF index.
[7] There are 1,621 FSAs in Canada.

NHS. The CCHS and the NHS are merged based on the residential FSA of the respondents which is only available in the master files.[8]

The CCHS is a cross-sectional survey that gathers health-related data of the Canadian population such as: health status, health care utilization and health determinants. In addition, the CCHS collects a wide range of socio-economic and demographic characteristics of the respondents. The large sample size of this survey provides the researcher with the opportunity to analyze the data within different groups of individuals or at a specific regional level. The CCHS 2013-2014 targets individuals living in private dwellings who are aged 12 years or older, and in total, includes information for 127,462 individuals living in the ten provinces and three territories.[9]

The Census of Population provides the most comprehensive demographic, social and economic statistical data on every person living in Canada. The census is conducted by Statistics Canada every five years, and all individuals are required to complete and return the census questionnaire. However, to maintain data confidentiality, researchers are provided by a sample of 1 in 5 extracted from Census of Population. Before 2011, the long mandatory questionnaires were sent to approximately 20 percent of respondents, however, in 2011, the long mandatory questionnaires were replaced by a voluntary long questionnaires that targeted to a random sample of 4.5 million households (NHS-2011).[10]

Dependent variables

The main dependent variable of this paper is a binary variable that takes the value of one if the respondent has ever been diagnosed with a sexually transmitted infection, zero otherwise.

[8] It is also important to mention that the publicly available 2013-2014 does not include one of the main dependent variables (number of sexual partners), and several explanatory variables are available only in a categorical format (e.g., age, income,...). In addition, the publicly available NHS only contains information on a 2.7% sample of respondents. Therefore, using the master file will provide us with a more precise estimate of the ethnic heterogeneity index.

[9] Approximately 40 percent of interviews were conducted in person and 60 percent over the phone. The response rates of households and single-persons are 75.9 and 87.3 percent, respectively. Data is collected directly from the survey respondents. Although proxy reporting is allowed in the CCHS, there is no case of proxy response in the sub-sample of the 2013-2014 used in this paper.

[10] The final participation rate for the NHS was 68.6%.

This question was asked to respondents aged 15 to 49 who answered yes or did not provide a valid answer to the question asking whether they have ever had sexual intercourse in their life.[11] Among all respondents who were between 15 and 49 years old, 39,830 of them are included in this study, as they provide a valid answer (yes or no) to the question related to the sexually transmitted infections.[12]

The secondary dependent variable of this paper is the respondent's number of different sexual partners in the 12 months preceding the survey. This question, that is only available in the master files of the CCHS, was asked to respondents aged 15 to 49 who have had sexual intercourse in the past 12 months. The results of both questions are presented in Table 3.1, part a.[13] It is worthwhile to mention that the reported number of sexual partner caps at four. In total, 3.6 percent of the individuals reported that they have had 4 partners or more in the 12 months preceding the survey. The average number of sexual partners is 1.17 when marital and common-law partners are included and 1.39 otherwise. The average prevalence of STDs within the sample is 0.082, increasing to 0.095 when excluding married individuals or those in a common-law relationship.

Figures 3.4 and 3.5 illustrate the relationship between FSA-level ethnic heterogeneity on the number of sexual partners and the prevalence of STDs, respectively. The average number of sexual partners and the prevalence of STDs are calculated for each quintile. The figures suggest that ethnic heterogeneity is positively correlated with both the number of sexual partners and the prevalence of STDs. Interestingly, the average number of partners and the average of prevalence of the STDs increases with HHI for the first four quintiles, which is consistent with the predictions of the model. However, both the average prevalence of STDs and the average number of partners drop for the fifth quintile. This can be explained

[11] In total, 50,785 out of 128,310 of the survey respondents were between 15 and 49 years old. Among them, 81.7 percent have had intercourse at least once in their life, while 11.9 percent have never been sexually inactive, and the rest (6.3 percent) did not provide a valid answer.

[12] This study excludes the data on 4,350 respondents that did not provide a valid answer to the main dependent variable of this study.

[13] All the regressions and descriptive analysis are weighted using the person weight variable: WTS_M. In the sample, each person is representative of a group of individuals with similar characteristics.

in two ways: first, living in a highly diversified neighborhood encourages individuals to break their social ties and communicate with members of other ethnic groups, and according the model, the ethnic heterogeneity increases the sexual infidelity only if communication and dissemination of information is limited across ethnic groups; second, some ethnic groups hold more conservative sexual attitude compared to the average population, and are less likely to provide accurate information about their sexual activities,[14] or get engaged in sexual partnerships (an individual would forgo the chance of finding a partner in that group due to the high marginal cost of searching in that group).

Explanatory variables

Based on the Employment Equity Act, the population of Canada has been classified into fourteen different ethnic groups: White, Aboriginal, South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean and Japanese, Other visible minorities, Multiple visible minorities. Detailed data on visible minority groups from the 2011 NHS were used to calculate ethnic heterogeneity (HHI), using Equation 3.1. The calculated HHI index is assigned to the individual records of CCHS based on the first three characters of the residential postal code of the respondents. Part b of Table 3.1 represents the summary statistic of the computed HHI.

To properly identify the effect of ethnic heterogeneity, this study controls for individual characteristics that might affect the spread of STDs. We control for the basic characteristic of individuals such as, age, gender, marital status (married, common-law, single), sexual orientation (heterosexual, homosexual, or bisexual), and migration status (Canadian born or foreign born). In addition, we include a binary variable, which takes the value of one if the respondents of the survey smokes cigarettes everyday, zero otherwise. We also control for the respondent's frequency of drinking alcoholic beverages by including a variable that lies between 0 and 7, where a higher number indicates more frequent consumption of alcoholic

[14] According to Kustec et al. (2016) and Harrington et al. (2001), members of certain religion and cultural groups and are more likely to underreport their sexual activities.

beverages.[15] Next, we include a dummy which is equal to one if the respondents hold a bachelor's degree or higher, and zero otherwise. Furthermore, we control for the occupation status of respondents by including a binary variable that takes the value of one if the respondents worked during the week preceding the survey, zero otherwise. Lastly, individuals are grouped according to 10,000 dollar intervals of household income:[16] the reference group corresponds to individuals whose reported yearly household income is between 0 and 10,000 dollars, while the final dummy accounts for the individuals with yearly household income exceeding 150,000 dollars. We include 14 dummies for the presence of the individuals in each category of household income.

Table 3.1, panel c displays summary statistics of the main controls variables used in this study.

3.5.2 Identification strategies

Quantifying the effect of FSA-level ethnic heterogeneity on the sexually transmitted disease is challenging. According to the Tiebout sorting model, while individuals are free to decide where to live and in what communities, their characteristics actively affect their selection. For instance, individuals living in families with kids tend to reside in neighborhoods with public schools. At the extreme, assuming only two neighborhoods, one neighborhood will end up composed entirely of families with kids while the other will only contain elderly and single individuals.[17] Similarly, in order to facilitate finding a sexual partner, individuals with more sexual desire might choose to reside in a neighborhood with a larger number of potential sexual partners (potential self-selection bias). This action, by itself, increases the chance of engaging in sexual activities, and hence increases the chances of getting infected.

[15] 0 for not drinking, 1 for drinking less than once a month, 2 for drinking once a month, 3 for drinking two or three times a month, 4 for drinking once a week, 5 for drinking two or three times a week, 6 for drinking four to six times a week, and 7 for drinking everyday.

[16] Household income data in the CCHS 2013-2014 is imputed, which means missing values due to lack of a valid answer to the household income are replaced using a nearest neighbour imputation method based on a modeled household income. In the sample of the CCHS 2013-2014 used for the purpose of this study, the data on house hold income of 10,213 (25.6 percent of) respondents were imputed.

[17] See Tiebout (1956).

Therefore, ethnic heterogeneity might be endogenous in this framework.

We address this issue of potential endogeneity in three different ways: first, by controlling for variables that might have an impact on the FSA-level value of ethnic diversity and at the same time on outcomes variables. For instance, controlling for education and income, which are indicators of individuals' ability. Second, by implementing an instrumental variables estimator using three sets of instrumental variables, including the 2006, 2001, and 1996 values of ethnic heterogeneity, average value of house prices,[18] and average value of household income as instrumental variables for the 2011 value of ethnic heterogeneity. Lastly, by conducting falsification tests and evaluating the effect of ethnic heterogeneity on non-transmitted diseases such as, asthma, arthritis, high blood pressure, heart disease, migraine, cancer, and diabetes.

Individual-specific controls

To estimate the implications of the model, we begin by examining the effect of ethnic diversity on the number of partners and on the likelihood of contracting sexually transmitted diseases. Specifically, we use Ordinary Least Square regression (OLS) to estimate our baseline model as follows:

$$y_{if} = \beta_0 + \beta_1 hhi_{if} + \gamma_e + \lambda_{cma} + \epsilon_{if}, \tag{3.2}$$

where y_{if} is the dependent variable. We run the regression twice: first, the dependent variable is the number of sexual partners of respondent i who resides in FSA f in the 12 months preceding the survey; second, the dependent variable is a binary variable which take the value of one if the respondent i, living in FSA f, has ever been diagnosed with an STD, and zero otherwise. The main variable of interest is the FSA-level value of ethnic heterogeneity, hhi_{if}, which is computed based on equation (3.1) and assigned to respondents

[18]House prices are driven based on a question that was asked to the respondent of the survey than owns their house; "If you were to sell your dwelling now, for how much would you expect to sell it".

according to their current residence. γ_e and λ_{cma} are ethnic and Canadian Metropolitan Areas (CMA) fixed effects,[19] respectively. Finally, ϵ_{if} is the error term. All error terms is this study are clustered at the FSA level (1,621 FSA groups).

The first approach to address endogeneity concerns is to control for variables that might have an impact on HHI and at the same time on the number of sexual partners and the prevalence of STDs. Importantly, we add proxy variables for ability which is unobservable. Following Altonji et al. (2005) and similar to Pongou (2009), we use the ratio of $\frac{\beta_F}{(\beta_R - \beta_F)}$ to measure how large the selection of unobservable variables need to be to explain away all the entire effect of FSA-level ethnic heterogeneity on the dependent variables, where β_F is the point estimate of HHI where a full set of control is used and and β_R represents the estimated coefficient of HHI when only a restricted set of observable controls is used. We re-estimate the baseline specification, as in 3.2, adding the extra individual specific controls as follows:

$$y_{if} = \beta_0 + \beta_1 hhi_{if} + C_{if}\Pi + A_{if}\Phi + \gamma_e + \lambda_{cma} + \epsilon_{if}, \tag{3.3}$$

where C_{if} is a vector of respondent i's basic personal characteristics, which includes age, gender, marital status, sexual orientation, migration status, and drinking and smoking habits. Second, we add a vector of respondent i's characteristics, A_{if}, which includes education, occupation, and household income that are proxy variables for ability (Ashenfelter and Rouse, 1998; Berry et al., 2006).

It is worthwhile to mention that, as both dependent variables of this paper are self-reported, it is likely that our estimates are biased downward in both scenarios. As sexual habits and attitudes are private information, depending on social and cultural norms, individuals might not reveal their real sexual behavioral patterns (this is known as the Social Desirability Bias). Kustec et al. (2016), Harrington et al. (2001), and Hser et al. (1999) show that individuals tend to underreport their positive STD status. Fenton et al. (2001), Brown and Sinclair (1999), and Kelly et al. (2013) show that the underreporting bias increases with

[19] 13 ethnic and 155 CMA dummies are included in the regressions.

the number of partners. Therefore, it is likely that the magnitude of the point estimates of the effect of ethnic heterogeneity on the number of partners and prevalence of STDs are likely to be the lower bound of the actual effects.

Instrumental variable approach

We tackle the potential endogeneity problem by instrumenting for the FSA-level value of ethnic heterogeneity using a set of variables from the past most recent three censuses (2006, 2001, and 1996). We require instruments that are correlated with the value of FSA-level ethnic diversity but uncorrelated with the unobservable characteristics of the ethnic groups that affect the prevalence of the STDs.

The first component of the IVs is the previous FSA-level values of ethnic diversity, which is inspired by Alesina et al. (1999). The identification assumption is that, the pervious value of HHI is highly correlated with the current ethnic ethnic heterogeneity, but unrelated to the current number of sexual partners or prevalence of STDs in a neighborhood. This reassures us that results of the paper are not contaminated by the recent endogenous migration.[20]

Each set of our IVs also include the average value of FSA-level house prices and household income, which are inspired by a large strand of economics literature that investigates the cause of migration into and out of a neighborhood. The results of both theoretical and empirical studies suggest that wages and house prices are the main factors affecting the household location decisions (Bover et al., 1989; Cameron and Muellbauer, 1998; Johnes and Hyclak, 1994; Strassmann, 2000; Plantinga et al., 2013). Individuals tend to relocate into neighborhoods with higher wages, while relatively higher house prices discourage migration into the neighborhood (Eichman et al., 2010). At the individual level of migration analysis, individuals are considered to be price-takers in either the labour or housing markets who have to accept the wages and house prices of the neighborhood they are moving into (Dahl and Sorenson, 2010; Kennan and Walker, 2011; Plantinga et al., 2013). In addition, it is unlikely

[20] We did not have earlier data on ethnic diversity to use as instrument, as the previous censuses do not report detailed data on visible minority groups.

that the previous average value of housing price and household income in a neighborhood is correlated with other factors that affect the prevalence of the STDs.

We use data from the 2006, 2001, and 1996 publicly available censuses to calculate the instrumental variables at the FSA-level.[21] The previous values of FSA-level ethnic heterogeneity, HHI, are calculated using Equation 3.1. The average value of FSA-level house price correspond to the average expected selling price of owned-occupied dwellings in each FSA.[22] Finally, the average value of FSA-level of household income refers to the total household income earned by all members of the family which has been divided by the square root of the family size of the household. This adjustment makes the household income comparable for different sizes of the family, as cost of living for additional member of a household is relatively lower than for an individual who lives alone.

Panel b of Table 3.1 illustrate the summary statistics of the IVs. As shown the average value of ethnic heterogeneity increased from 23.52 percent in 1996 to 33.75 percent in 2011. Panel D of Table 3.1 displays the average value of dwellings and household income for in 2011, 2006, 2001, and 1996. The average value of dwelling (the average value of household income) increased by 68 (113) percent between 1996 and 2011.

Identification checks

In the recent years, a growing body of economics literature has employed "tests of design" to gauge the plausibility of models or identification strategies Hartman et al. (2015). Placebo tests are widely used to evaluate whether the apparent casual effect have to do with the research design, and not with the presence and effects of hypothesized causal mechanisms.

Following Pongou (2009), we use placebo diseases to examine the effect of ethnic heterogeneity on a set of alternative dependent outcomes. Placebo diseases are seven different non-transmitted diseases.[23] If the intervention shows that the ethnic heterogeneity is corre-

[21] A 20 percent sample representative of the whole population has been used in the census public data.
[22] Tenant-occupied, band houses, and farm dwellings are not included.
[23] Including asthma, arthritis, high blood pressure, heart disease, migraine, cancer, and diabetes.

lated with the series of non-transmitted diseases then the validity of the research design is called into question.

3.6 Results

3.6.1 Main

The main results are summarized in Tables 3.2 and 3.3, showing the effect of ethnic heterogeneity on the number of partners and the prevalence of sexually transmitted diseases, respectively using different sets of controls.

Table 3.2 reports the estimated effect of the FSA-level ethnic diversity on the number of sexual partners in the 12 months preceding the survey. Column 1 shows that the estimated coefficient of ethnic heterogeneity in 0.159, when only CMA and ethnic fixed effects are used. This result indicates that a one standard deviation increase in the residential neighbourhood of individual leads to an increase in the number of his/her sexual partners equal to 0.0364 units. After adding age and gender controls the point estimate drops to 0.123 (column 2). We re-run the estimation while adding a large set of individual-level controls. Column 3 represents that the point estimate falls to 0.088 after controlling for marital status, sexual orientation, smoking and drinking habits, a binary indicator of being an immigrant. The estimated coefficient of the effect of HHI on the number of sexual partners increases slightly to 0.094 when we control for education, occupation status, and household income (column 4). To investigate the effect of type of the residential area (urban vs rural), we re-run the same regression twice; first, we add a dummy that take the value of one if the individual is residing in an urban area, otherwise zero and second, we re-run the estimates for those individuals who are residing in urban area, separately. The results of these practice are shown in columns 5 and 6, respectively. In both cases the point estimates are significant and the magnitudes change slightly. Finally, column 7 displays the effect of ethnic diversity on the number of sexual partners for those individuals who are in a committed relationship (married or common-law partners), as we can see the point estimate is almost half of the point estimate with the same specification (column 4). These results are not surprising, considering that the dependent variable is the number of sexual partners in the 12 months preceding the survey, it is highly expected that individuals in a committed relationships do not revel the true number of sexual partners in case of infidelity.

Table 3.3 displays the estimates of the effect of the FSA-level ethnic heterogeneity on the prevalence of STDs. Column 1 displays the results for the basic specification in which we only include CMA and ethnic fixed effects (the point estimate of HHI is equal to 0.0444). In this scenario a one standard deviation increase in the ethnic diversity of the residential neighbourhood of a sexually active individual will increase the chance of getting infected by 1.015 percent. Column 2 illustrates that the point estimate of HHI increases to 4.79 percent when we add age and gender controls. The probability of getting infected increases to 5.29 when we add the second set of individual controls, including marital status, sexual orientation, smoking and drinking habits, and immigrant status (column 3). Finally, in column 4, if we additionally control for education, occupation, and household income the likelihood of getting infected declines to 4.71 percent. The magnitude and significance of the HHI point estimates are notably stable, considering that all the controls that are added to the model are powerful and statistically significant. Columns 5 and 6, similar to table 3.2, display the results of the same specification controlling for the type of residential area (urban vs rural). In column 5 we add a dummy that take the value of one if the individual is residing in an urban area, otherwise zero and in column 6 , we re-run the regression for those individuals who are residing in urban area, separately. In both cases the significance of the point estimates remain intact and the change in the magnitudes are negligible. Column 7 represents the effect of ethnic diversity on the prevalence of STDs within individuals who are in a committed relationship (married or common-law partners), as we can see the magnitude of the point estimate is bigger and it remains significant. This is suggestive that the results of this paper is not entirely driven by individuals who are out of a committed relationship.

As discussed before, unobservable variables such as ability might affect HHI and at the same time number of sexual partner and probability of getting infected with an STD. Using the ratio of $\frac{\beta_F}{(\beta_R - \beta_F)}$, we measure how large the selection of unobservable variables need to be to explain away all the entire effect of FSA-level ethnic heterogeneity on the dependent variables. Column 4 of Tables 3.2 and 3.3 illustrate the point estimate of HHI where a full set of control is used (β_F), while column 2 of Tables 3.2 and 3.3 represents the estimated coefficient of HHI when only a restricted set of observable controls is used (β_R). According to our estimates, the selection of unobservable variables would have to be 3.30 and 58.88 times greater than the selection of observable variables to attribute the effect of HHI on the number of partners and the probability of getting infected with an STD, respectively. Both computed ratios are significantly greater than one, which suggests that it is not likely that the entire estimated effects are driven by the selection of the unobservables.

We grouped individuals according to their ethnic group, migration status, and gender, and estimate the effect of FSA-level ethnic heterogeneity on the number of sexual partners and the prevalence of STDs within each group separately. In total, we have 10 different groups: whites, white non-immigrants, white immigrants, aboriginals, visible minorities,[24] visible minorities non-immigrants, visible minority immigrants, immigrants, males, and females. Tables 3.4 and 3.5 report the results of this practice. The effect of ethnic diversity on both STDs and number of sexual partners become insignificant when we only include immigrants in our regressions (columns 3, 7, and 10). It is predictable that we see little or no effect among immigrants from different ethnic groups. First, it might be more difficult for immigrants to find a sexual partner due to cultural and language barriers. Second, immigrants of certain countries might have more conservative sexual behaviours (i.e., Muslim immigrants). Third, the government of Canada requires all individuals who apply to become a permanent resident of Canada and applicants of temporary resident visa who lived in certain countries or territories[25] more than six months in the preceding year before entering Canada to pass

[24]Everyone except whites and aboriginals.

[25]146 out of 279 countries or territories. For a full list of countries visit `http://www.cic.gc.ca/english/`

a medical exam. The exam includes an HIV/ADIS and Rapid Plasma Reagin (RPR), which is useful in diagnosing syphilis infection. Individuals who tests positive are banned from entering Canada.

The estimated coefficient of HHI on the prevalance of STDs is bigger when we only include females compared to when only men are included (columns 8 and 9). When caught cheating, women are often more severely punished than men, so they try harder to hide their infidelity. It is therefore not surprising that the effect of ethnic heterogeneity is larger within the female cohort. This result is aligned with the prediction of the model, and also consistent with the findings of Cahu et al. (2014) and Pongou (2009).

3.6.2 Instrumental variables

The results of IV estimates are illustrated in Tables 3.7 and 3.8 including a full set of controls. In order to make sure that the exclusion restrictions are satisfied, we only report the IV estimation results after adding the full set of controls, similar to the specification of the last column of Tables 3.2 and 3.3.

Table 3.7 reports the IV estimates of ethnic heterogeneity on the number of sexual partners, and Table 3.8 shows the IV estimates of ethnic diversity on the prevalence of STDs. The full set of control variables were included in these regressions: the individual controls, ethnic controls, and CMA controls. Columns 2, 3, and 4 of Tables 3.7 and 3.8 report the IV estimates in which we use the previous value of FSA-level ethnic heterogeneity, the average value of house prices, and the average value of household income in the 2006, 2001, and 1996 censuses, respectively. The second-stage estimates show a positive and highly significant effect of ethnic heterogeneity on the number of sexual partners and the prevalence of STDs. Furthermore, the magnitude of point estimates are close to the OLS estimates. These results suggest that the presence of potential omitted variables is not strongly biasing the OLS estimates.

information/medical/dcl.asp.

As we mentioned before, the dependent variable, STD, is equal to one if the respondent has ever been diagnosed with a STD, zero otherwise. As we do not have information on the exact time of infection, we check the robustness of the model by excluding the individuals who have been sexually active before 1996 (whom have been sexually active for more than 17 years), and using the 1996 FSA-level value of FSA-level ethnic heterogeneity, average value of house prices, and Household income as instrumental variables. Column 6 of Table 3.8 reports the results of this exercise, the point estimate of the ethnic heterogeneity coefficient on the prevalence of STDs remains significant and magnitude is slightly higher than the OLD estimates.

3.6.3 Falsification test: The effect of ethnic diversity of the non-transmissible diseases

The results of the paper suggest that an increase in ethnic diversity in a FSA increases the prevalence of STDs. We test the validity of the casual inference by replacing the dependent variable (STD) with seven different non-transmitted diseases, outcomes that should not be affected by HHI: asthma, arthritis, high blood pressure, heart disease, migraine, cancer, and diabetes. In all regression, the dependent variable is a dummy that takes the value of one if the respondent of the survey is suffering from the marked disease, zero otherwise.

Table 3.9 illustrates the results of this exercise. As shown in column 1 to 7 all the coefficients are statistically insignificant. The absolute value of the point estimates are smaller than in the original model, and the signs of the coefficients are mixed. These findings suggest that it is not likely that the value of ethnic heterogeneity has an impact on the prevalence of non-infectious diseases, which indicate that our mechanism passes the falsification test. These results are consistent with the findings of the falsification test conducted by Pongou (2009), which shows that the value of ethnic diversity does not affect the prevalence of infectious diseases which are not sexually transmitted.

3.7 Conclusion

This study contributes to a growing body of literature that focuses on understanding the dynamic of the formation of sexual networks as an essential tool for identifying the determinants of the spread of STDs. We apply sexual network analysis to explain how ethnic diversity affects the formation of sexual network and hence, the spread of sexually transmitted diseases in a multicultural framework. Optimizing individuals ethnically diversify their sexual partners in order to hide their infidelity behind the anonymity nurtured by the slow dissemination of information across different ethnic groups within a society. This optimizing behavior facilitates the spread of STDs by increasing sexual infidelity and broadening sexual networks among ethnic groups.

We employ a detailed data set of 39,830 sexually active individuals aged between 15 to 49, obtained from the 2013-2014 CCHS, to empirically test the implications of the model. The results confirms that individuals living in the highly ethnically heterogenous neighborhoods have more sexual partners and are more susceptible to STDs compared to residents of ethnically homogeneous neighborhoods.

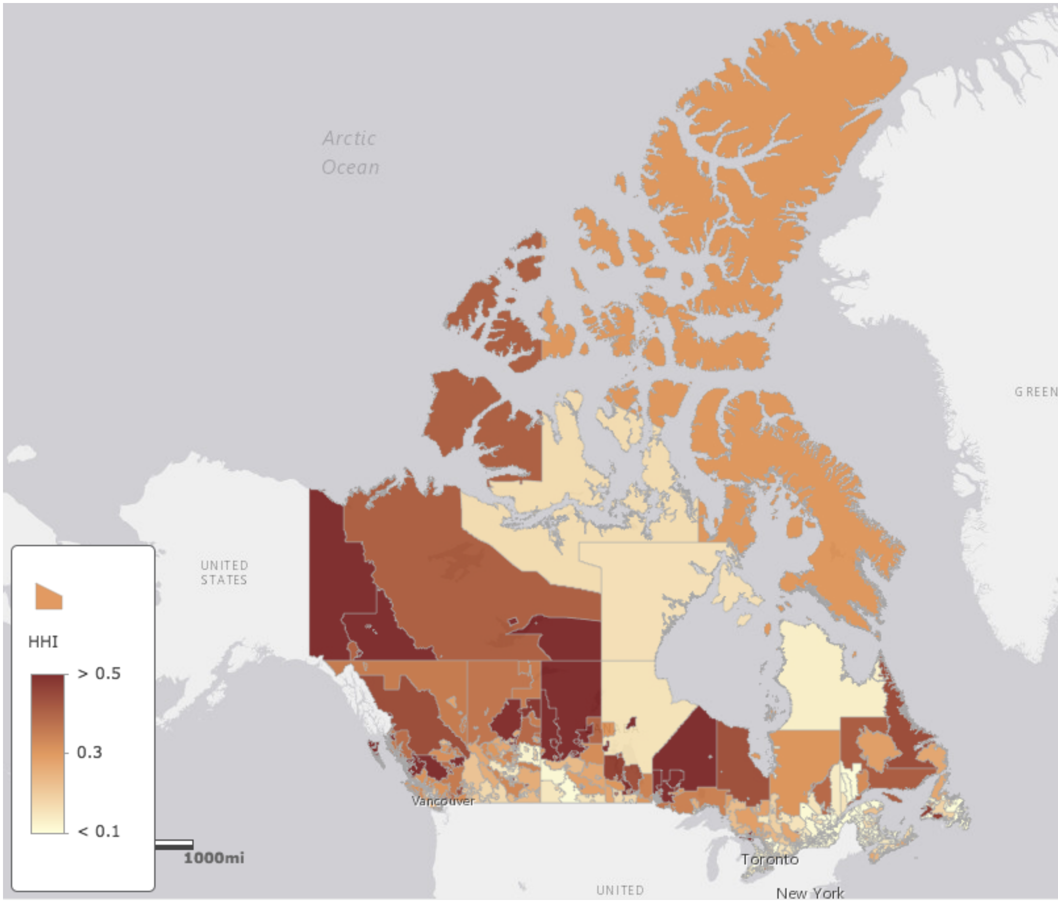
In order to ensure that the relationship between HHI and outcomes is causal we use three strategies. First, using the method developed by Altonji et al. (2005) we show that the selection on unobservable variables needs to be 58.88 (3.30) times greater to attribute the entire effect of ethnic diversity on the prevalence of STDs (the total number of sexual partners). Second, we use the 2006, 2001, and 1996 values of HHI, average house prices, and average household income as instrumental variables for HHI. The IV estimates results confirm the positive effect of HHI on outcomes. Finally, the results of the placebo test show that ethnic heterogeneity does not significantly affect the prevalence of non-communicable diseases in our sample.

Appendix

Figure 3.1: Forward Sortation Area*

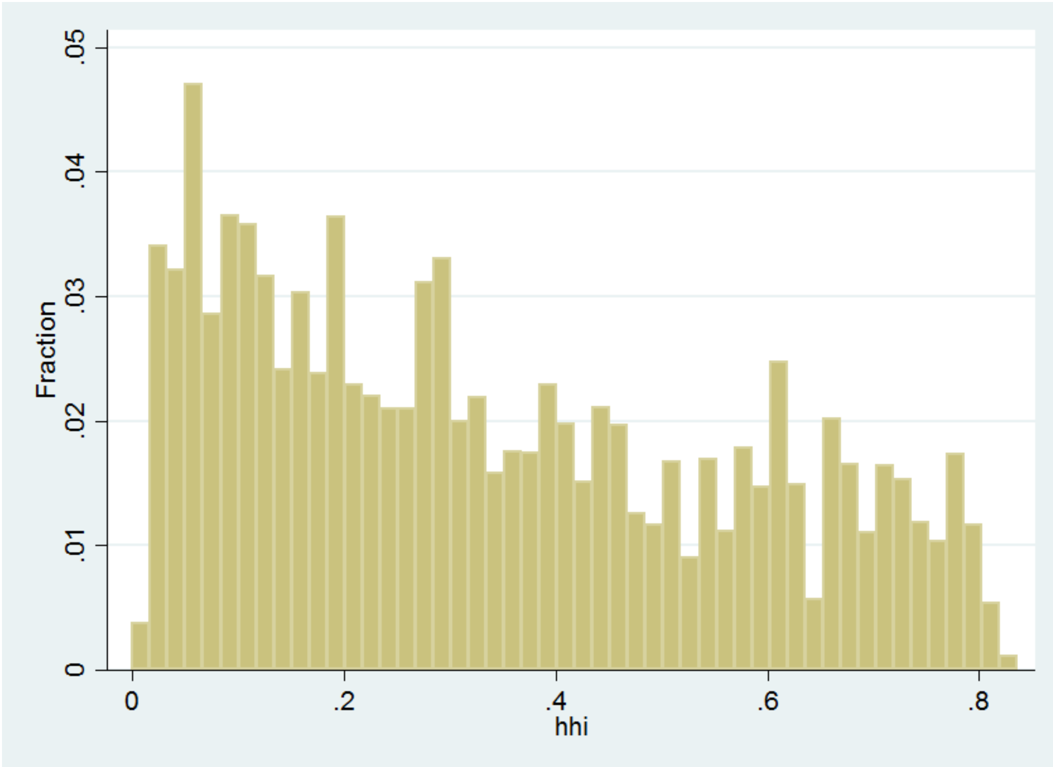
Source: Innovation, Science and Economic Development Canada (ISED), available at `https://www.ic.gc.ca/eic/site/bsf-osb.nsf/eng/br03396.html`

Figure 3.2: Ethnic diversity*



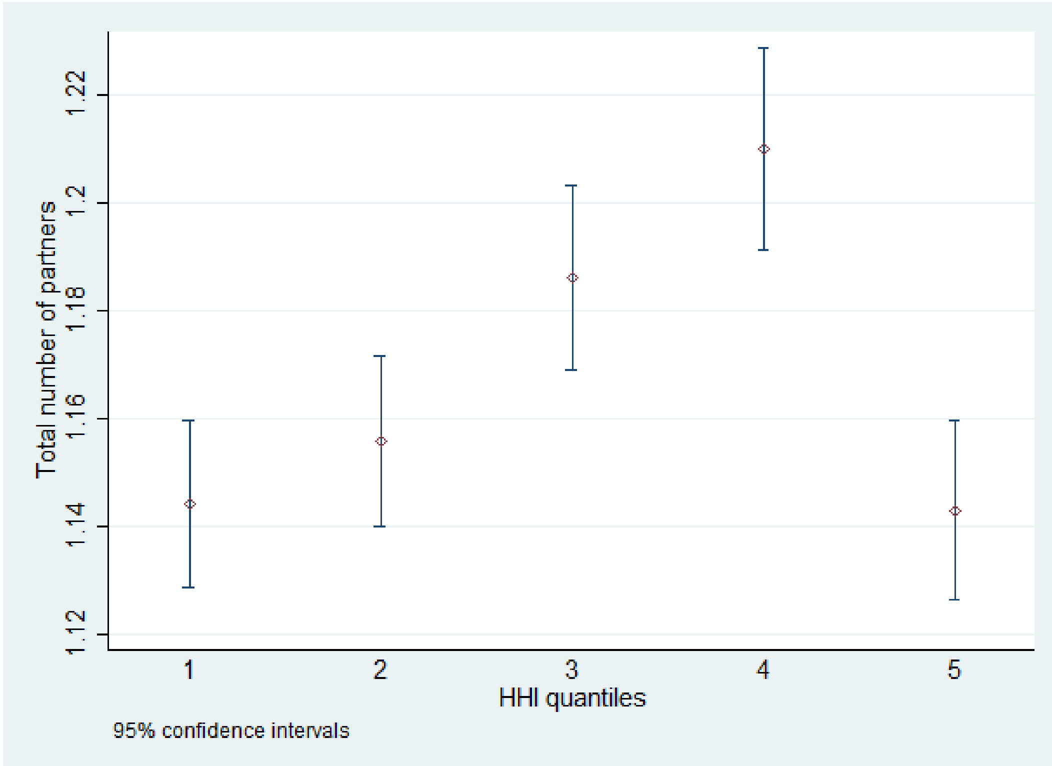
Source: Statistics Canada.
The map is created using ArcGIS Online mapping tools.

Figure 3.3: Distribution of ethnic diversity index (HHI)*



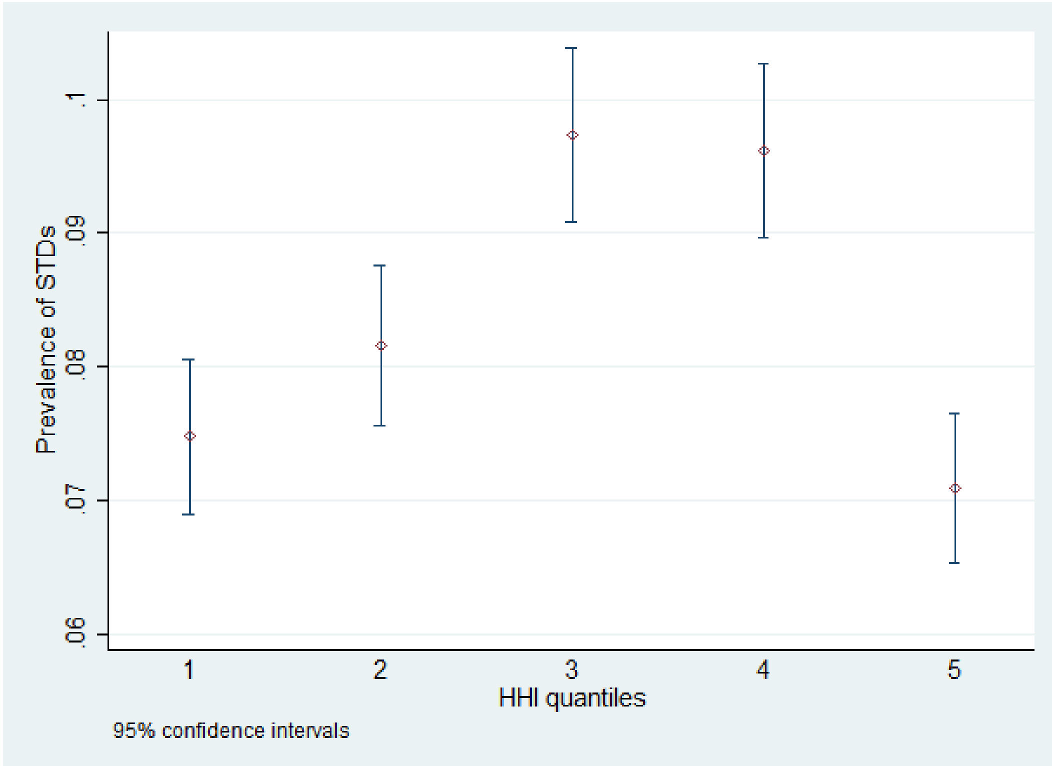
Source: Statistics Canada.

Figure 3.4: Effect of FSA-level ethnic heterogeneity on the number of sexual partners*



Note: Error bars show 95 percent confidence interval around the mean.

Figure 3.5: Effect of FSA-level ethnic heterogeneity on the prevalence of STDs*



Note: Error bars show 95 percent confidence interval around the mean.

Table 3.1: Summary statistics*

Variable	Observations	Mean	Std. Dev.
Part a			
STDs	39,830	0.08251	0.2751
STDs (excluding marital & common law partners)	18,435	0.0951	0.2933
STDs (just marital & common law partners)	21,395	0.0738	0.2615
STDs (males)	18,320	0.0638	0.2445
STDs (females)	21,510	0.1016	0.3022
# of sexual partners	39,795	1.1651	0.7624
# of sexual partners (excluding marital & common law partners)	18,405	1.3861	1.0984
# of sexual partners (Just marital & common law partners)	21,390	1.0125	0.3043
# of sexual partners (males)	18,320	1.2289	0.8461
# of sexual partners (females)	21,510	1.0999	0.6599
1 sexual partner	39,795	0.784	0.4115
2 sexual partner	39,795	0.0674	0.2507
3 sexual partner	39,795	0.0329	0.1783
4 or more sexual partner	39,795	0.036	0.1882
Part b			
HHI	39,830	0.3375	0.2287
HHI 2006	39,755	0.2989	0.2244
HHI 2001	39,310	0.2601	0.2154
HHI 1996	38,040	0.2352	0.2028
Part c			
Male	39,830	0.5059	0.4500
Age	39,830	33.7789	9.2288
Age of first intercourse	39,830	17.465	3.2318
Married	39,830	0.3792	0.4852
Common-law	39,830	0.1581	0.3648
Immigrant	39,830	0.2282	0.4197

Notes: The summary statistics are weighted.

(continued on the next page)

Summary statistics (continued from previous page)*

Variable	**Observations**	**Mean**	**Std. Dev.**
Part c (continued from previous page)			
Bisex	39,830	0.0143	0.1187
Homosex	39,830	0.0176	0.1316
Alcohol frequency	39,830	2.6145	1.5290
Daily smoking	39,830	0.1931	0.3947
Education	39,830	0.2926	0.4550
Worked last week	39,830	0.7552	0.4299
Household Income	39,830	8.62211	4.0345
Part d			
Mean value of dwellings 2011	39,830	267,988.8	151,385.4
Mean households income 2011	39,830	94,479.33	29,486.63
Mean value of dwellings 2006	39,755	208,276.2	121,099.9
Mean households income 2006	39,755	86,177.97	37,350.67
Mean value of dwellings 2001	39,310	172,125.5	82,954.7
Mean households income 2001	39,310	61,555.59	19,146.85
Mean value of dwellings 1996	38,040	159,444.8	80,434.67
Mean households income 1996	38,040	44,329.03	13,431.42

Notes: The summary statistics are weighted.

Table 3.2: The effect of ethnic heterogeneity on the number of partners*

Dependent Variable:	Number of Sexual Partners						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HHI	0.159***	0.123**	0.0883*	0.0944**	0.0816*	0.0892*	0.0472
	(0.0518)	(0.0514)	(0.0464)	(0.0465)	(0.0470)	(0.0528)	(0.0339)
Age		-0.0167***	-0.0111***	-0.0112***	-0.0112***	-0.0116***	-0.0032***
		(0.0007)	(0.0008)	(0.0008)	(0.0010)	(0.0008)	(0.0006)
Gender		0.124***	0.0711***	0.0645***	0.0645***	0.0669***	0.0234***
		(0.0121)	(0.0118)	(0.0118)	(0.0118)	(0.0135)	(0.00696)
Common-law			-0.289***	-0.291***	-0.290***	-0.303***	0.0335***
			(0.0178)	(0.0180)	(0.0180)	(0.0206)	(0.0118)
Married			-0.224***	-0.224***	-0.227***	-0.224***	
			(0.0170)	(0.0177)	(0.0177)	(0.0197)	
Homosexual			0.344***	0.343***	0.343***	0.376***	0.233
			(0.104)	(0.104)	(0.104)	(0.111)	(0.143)
Bisexual			0.401***	0.408***	0.407***	0.444***	0.190**
			(0.0665)	(0.0661)	(0.0661)	(0.0723)	(0.0873)
Daily Smoking			0.0973***	0.0966***	0.0965***	0.100***	0.0239*
			(0.0182)	(0.0186)	(0.0186)	(0.0220)	(0.0141)
Immigrant			0.00949	0.0165	0.0161	0.0227	-0.00458
			(0.0242)	(0.0245)	(0.0245)	(0.0262)	(0.0181)
Alcohol Frequency			0.0730***	0.0720***	0.0718***	0.0735***	0.0087***
			(0.0042)	(0.0043)	(0.0049)	(0.0043)	(0.00254)
Education	N	N	N	Y	Y	Y	Y
Worked Last Week	N	N	N	Y	Y	Y	Y
HH Income	N	N	N	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y	Y	Y	Y
CMA	Y	Y	Y	Y	Y	Y	Y
Urban	N	N	N	N	Y	N	N
Sample Restriction	NA	NA	NA	NA	NA	Urban-Area	Married & Common-law
Observations	39,795	39,795	39,795	39,795	39,795	29,870	21,390

Notes: The dependent variable is the number of respondent's different sexual partners in the 12 months preceding the interview. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.3: The effect of ethnic hetherogeneity on the prevalence of STDs*

Dependent Variable:	STD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HHI	0.0444**	0.0479***	0.0529***	0.0471***	0.0397**	0.0439**	0.0544**
	(0.0184)	(0.0184)	(0.0176)	(0.0176)	(0.0179)	(0.0203)	(0.0219)
Age		0.0011***	0.0023***	0.0023***	0.0023***	0.0026***	0.0007
		(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Gender		-0.0384***	-0.0461***	-0.0433***	-0.0432***	-0.0471***	-0.0355***
		(0.0047)	(0.0047)	(0.0047)	(0.00471)	(0.00545)	(0.00623)
common-law			0.00796	0.0115	0.0116	0.0104	
			(0.0078)	(0.0078)	(0.00779)	(0.00930)	
Married			-0.0382***	-0.0331***	-0.0327***	-0.0368***	-0.0413***
			(0.0062)	(0.0061)	(0.00611)	(0.00692)	(0.00831)
Homosexual			0.122***	0.122***	0.122***	0.123***	0.0986*
			(0.0316)	(0.0315)	(0.0315)	(0.0341)	(0.0575)
Bisexual			0.138***	0.133***	0.133***	0.136***	0.0996**
			(0.0285)	(0.0283)	(0.0283)	(0.0314)	(0.0399)
Daily Smoking			0.0436***	0.0404***	0.0404***	0.0386***	0.0431***
			(0.0067)	(0.0068)	(0.00677)	(0.00777)	(0.0105)
Immigrant			-0.0429***	-0.0465***	-0.0467***	-0.0487***	-0.0483***
			(0.007)	(0.0069)	(0.0069)	(0.0073)	(0.0084)
Alcohol Frequency			0.007***	0.0085***	0.0084***	0.0083***	0.0057***
			(0.0015)	(0.0016)	(0.0016)	(0.0018)	(0.0021)
Education	N	N	N	Y	Y	Y	Y
Worked Last Week	N	N	N	Y	Y	Y	Y
HH Income	N	N	N	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y	Y	Y	Y
CMA	Y	Y	Y	Y	Y	Y	Y
Urban	N	N	N	N	Y	N	N
Sample Restriction	NA	NA	NA	NA	NA	Urban-Area	Married & Common-law
Observations	39,830	39,830	39,830	39,830	39,830	29,900	21,400

Notes: The dependent variable is a dummy variable that takes the value of one if the respondent of the survey has been diagnosed with a sexually transmitted disease, zero otherwise. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.4: Coefficients on Ethnic Heterogeneity for the number of partners within different groups*

Dependent Variable:	STD									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample:	Whites	White non-immigrants	White immigrants	Aboriginals	Visible Minorities	Visible Minority non-immigrants	Visible Minority immigrants	Immigrants	Males	Females
HHI	0.109**	0.110**	0.00748	-0.00865	0.0671	-0.126	0.139	0.105	0.186**	-0.0173
	(0.0515)	(0.0554)	(0.153)	(0.182)	(0.124)	(0.257)	(0.142)	(0.110)	(0.0789)	(0.0486)
Individual-specific controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
CMA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	31,890	30,110	1,780	3,170	4,735	1,370	3,370	5,160	18,295	21,500

Notes: The dependent variable is the number of respondent's different sexual partners in the 12 months preceding the interview. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. Columns 1, 2, and 3 display the results of regression including the sample of white individuals. Column 2 only includes Canada-born white individuals born in Canada; Column 3 only includes the foreign-born white individuals. Only aboriginal people are included in the estimation of Column 4. Columns 5, 6, and 7 display the results of regression including the sample of visible minority individuals. Column 6 only includes Canada-born visible minorities, Column 7 only includes the foreign-born visible minority individuals. Only foreign-born individuals are included in the estimation of Column 8. Column 9 and 10 shows the results of the regressions including only male and female, respectively. All regressions include a constant, and a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.5: Coefficients on ethnic heterogeneity for the prevalence of STDs within different groups*

Dependent Variable:	STD									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample:	Whites	White non-immigrants	White immigrants	Aboriginals	Visible Minorities	Visible Minority non-immigrants	Visible Minority immigrants	Immigrants	Males	Females
HHI	0.0438^{**}	0.0506^{**}	-0.00921	0.0830	0.0682^{**}	0.122^{*}	0.0601	0.0472	0.0189	0.0764^{***}
	(0.0212)	(0.0229)	(0.0445)	(0.0814)	(0.0330)	(0.0632)	(0.0377)	(0.0311)	(0.0229)	(0.0260)
Individual-specific controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
CMA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	31,915	30,135	1,780	3,180	4,740	1,370	3,370	5,160	18,320	21,510

Notes: The dependent variable is a dummy variable that takes the value of one if the respondent of the survey has been diagnosed with a sexually transmitted disease, zero otherwise. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. Columns 1, 2, and 3 display the results of regression including the sample of white individuals. Column 2 only includes Canada-born white individuals born in Canada; Column 3 only includes the foreign-born white individuals. Only aboriginal people are included in the estimation of Column 4. Columns 5, 6, and 7 display the results of regression including the sample of visible minority individuals. Column 6 only includes Canada-born visible minorities, Column 7 only includes the foreign-born visible minority individuals. Only foreign-born individuals are included in the estimation of Column 8. Column 9 and 10 shows the results of the regressions including only male and female, respectively. All regressions include a constant, and a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.6: First-Stage Estimates*

	(1) HHI	(2) HHI	(3) HHI
2006 HHI	0.975***		
	(0.0108)		
2006 House Prices	0.0407		
	(0.0185)		
2006 HH Income	-0.00752		
	(0.0394)		
2001 HHI		0.924***	
		(0.0318)	
2001 House Price		-0.108*	
		(0.0597)	
2001 HH Income		0.718***	
		(0.205)	
1996 HHI			0.886***
			(0.0324)
1996 House Prices			-0.0257
			(0.0685)
1996 HH Income			0.805***
			(0.29)
Individual-specific controls	Y	Y	Y
Ethnic fixed effects	Y	Y	Y
CMA fixed effects	Y	Y	Y
Observations	39,755	39,310	38,040

Notes: HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. We instrument HHI using three different sets of instrumental variables; 2006 FSA-level value of HHI, average house prices, and average household income (Column 1), 2006 FSA-level value of HHI, average house prices, and average household income (Column 2), and 1996 FSA-level value of HHI, average house prices, and average household income (Column 3). All the first stage estimates include a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.7: IV estimates of the effect of ethnic heterogeneity on the number of partners*

Dependent Variable:	Number of Sexual Partners			
	(1)	(2)	(3)	(4)
	OLS	IV	IV	IV
		2006 HHI, house prices, & HH income	2001 HHI, house prices, & HH income	1996 HHI, house prices, &HH income
HHI	0.0944**	0.0970**	0.0988*	0.119**
	(0.0465)	(0.0484)	(0.0532)	(0.0572)
Individual-specific controls	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y
CMA	Y	Y	Y	Y
1st stage F-stat		2957.1	375.5	254.2
Observations	39,795	39,720	39,275	38,010

Notes: The dependent variable is the number of respondent's different sexual partners in the 12 months preceding the interview. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. In the columns marked IV, we instrument HHI using three different sets of instrumental variables; 2006 FSA-level value of HHI, average house prices, and average household income (Column 2), 2006 FSA-level value of HHI, average house prices, and average household income (Column 3), and 1996 FSA-level value of HHI, average house prices, and average household income (Column 4). All regressions include a constant, and a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.8: IV estimates of the effect of ethnic heterogeneity on the prevalence of STDs*

Dependent Variable:	STD				
	(1)	(2)	(3)	(4)	(5)
	OLS	IV	IV	IV	IV
		2006 house prices, HH income, & HHI	2001 house prices, HH income, &HHI	1996 house prices, HH income, & HHI	1996 house prices, HH income, & HHI
HHI	0.0471***	0.0493***	0.0543***	0.0640***	0.0497*
	(0.0176)	(0.0181)	(0.02)	(0.0227)	(0.029)
Individual-specific controls	Y	Y	Y	Y	Y
Ethnic fixed effects	Y	Y	Y	Y	Y
CMA fixed effects	Y	Y	Y	Y	Y
Years of sexual activity≤17	N	N	N	N	Y
1st stage F-stat		2957.7	375.6	254.3	225.8
Observations	39,830	39,755	39,310	38,040	20,170

Notes: The dependent variable is a dummy variable that takes the value of one if the respondent of the survey has been diagnosed with a sexually transmitted disease, zero otherwise. HHI is an index of ethnic heterogeneity that measures the probability that two different randomly chosen individuals from a population are from different ethnic groups, and is calculated at the FSA-level using Equation 3.1. In the columns marked IV, we instrument HHI using three different sets of instrumental variables; 2006 FSA-level value of HHI, average house prices, and average household income (Column 2), 2006 FSA-level value of HHI, average house prices, and average household income (Column 3), and 1996 FSA-level value of HHI, average house prices, and average household income (Column 4 and 5). Column 5 displays the results of regressions only including the sample of individuals who have been sexually active for less than 17 years. All regressions include a constant, and a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 3.9: Falsification: Coefficients on ethnic heterogeneity for the prevalence of non-transmitted diseases*

	(1) Asthma	(2) Arthritis	(3) Blood Pressure	(4) Heart Disease	(5) Migraine	(6) Cancer	(7) Diabetes
HHI	-0.0181	0.0026	0.0240	-0.0036	0.0153	-0.0081	0.0066
	(0.0176)	(0.0162)	(0.0177)	(0.0058)	(0.0226)	(0.0103)	(0.0095)
Individual-specific controls	Y	Y	Y	Y	Y	Y	Y
Ethnic	Y	Y	Y	Y	Y	Y	Y
CMA	Y	Y	Y	Y	Y	Y	Y
Observations	39,830	39,830	39,830	39,830	39,830	39,830	39,830

Notes: Dependent variables are binary variables which take the value of one if the respondent of the survey has the marked disease, zero otherwise. All regressions include a constant, and a full set of controls including; age, gender, marital status, sexual orientation, smoking and alcohol habits, immigration status, level of education, occupation status, household income, and ethnic and CMA fixed effects. All regressions include a constant, and weighted clustered standard errors (at the FSA level) are in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%.

*Data of Figures 3.1, 3.2, 3.3, 3.4, and 3.5, and Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, and 3.9 include information copied with permission from Canada Post Corporation. The attached custom data at the FSA level is provided for use in accordance with the terms and conditions of the Statistics Canada Open License Agreement.

Bibliography

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Allen, F. and Gale, D. (2000). *Comparing financial systems*. MIT press.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1):151–184.

Anderson, R. M. (1999). Transmission dynamics of sexually transmitted infections. *Sexually Transmitted Diseases*, pages 25–37.

Angelini, P. and Cetorelli, N. (2003). The effects of regulatory reform on competition in the banking industry. *Journal of Money, Credit and Banking*, 35(5):663–684.

Ashenfelter, O. and Rouse, C. (1998). Income, schooling, and ability: Evidence from a new sample of identical twins. *The Quarterly Journal of Economics*, 113(1):253–284.

Bajaj, S., Ramayanam, S., Enebeli, S., Nsohtabien, H., Andkhoie, M., Yaghoubi, M., Gall, S., Szafron, M., and Farag, M. (2017). Risk factors for sexually transmitted diseases in Canada and provincial variations. *Social Medicine*, 11(2):62–69.

Baron, D. P. (1989). Service-induced campaign contributions and the electoral equilibrium. *The Quarterly Journal of Economics*, 104(1):45–72.

Baron, D. P. (1994). Electoral competition with informed and uninformed voters. *American Political Science Review*, 88(1):33–47.

Becerra, O., Cavallo, E., and Scartascini, C. (2012). The politics of financial development: The role of interest groups and government capabilities. *Journal of Banking & Finance*, 36(3):626–643.

Benston, G. J., Hanweck, G. A., and Humphrey, D. B. (1982). Scale economies in banking: A restructuring and reassessment. *Journal of Money, Credit and Banking*, 14(4):435–456.

Berg, S. A. and Kim, M. (1994). Oligopolistic interdependence and the structure of production in banking: An empirical evaluation. *Journal of Money, Credit and Banking*, 26(2):309–322.

Berger, A. N., Klapper, L. F., and Turk-Ariss, R. (2009). Bank competition and financial stability. *Journal of Financial Services Research*, 35(2):99–118.

Bernheim, B. D. and Whinston, M. D. (1986). Menu auctions, resource allocation, and economic influence. *Quarterly Journal of Economics*, 101(1):1–31.

Berry, C. M., Gruys, M. L., and Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91(3):696–705.

Bertrand, M., Kramarz, F., Schoar, A., and Thesmar, D. (2004). Politically connected CEOs and corporate outcomes: Evidence from France. Technical report, University of Chicago.

Bikker, J. A. and Haaf, K. (2002). Measures of competition and concentration in the banking industry: A review of the literature. *Economic & Financial Modelling*, 9(2):53–98.

Blais, A. (2000). *To vote or not to vote?: The merits and limits of rational choice theory.* University of Pittsburgh Press.

Bolt, W. and Humphrey, D. (2012). A frontier measure of U.S. banking competition. Technical report, DNB Working Paper.

Bönisch, P., Geys, B., and Michelsen, C. (2015). David and Goliath in the poll booth: Group size, voting power and voter turnout. Technical report, DIW Berlin, German Institute for Economic Research.

Bover, O., Muellbauer, J., and Murphy, A. (1989). Housing, wages and uk labour markets. *Oxford Bulletin of Economics and Statistics*, 51(2):97–136.

Braun, M. and Raddatz, C. E. (2009). Banking on politics. Technical report, World Bank.

Brown, N. R. and Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research*, 36(3):292–297.

Cahu, P., Fall, F., and Pongou, R. (2014). Beauty, polygyny and fertility: Theory and evidence. Technical report, Documents de travail du Centre d'Economie de la Sorbonne 2014.78 - ISSN: 1955-611X. 2014.

Cameron, G. and Muellbauer, J. (1998). The housing market and regional commuting and migration choices. *Scottish Journal of Political Economy*, 45(4):420–446.

Campbell, A., Gurin, G., and Miller, W. E. (1954). *The voter decides.* Row, Peterson, and Co.

Carlton, D. W. and Perloff, J. M. (1990). *Modern industrial organization.* Scott, Foresman/Little, Brown Higher Education.

Chen, H., Parsley, D., and Yang, Y. W. (2015). Corporate lobbying and firm performance. *Journal of Business Finance and Accounting*, 42(3-4):444–481.

Chowell, G. and Hyman, J. M. (2016). *Mathematical and statistical modelling for emerging and re-emerging infectious diseases*. Springer.

Claessens, S., Feijen, E., and Laeven, L. (2008). Political connections and preferential access to finance: The role of campaign contributions. *Journal of Financial Economics*, 88(3):554–580.

Claessens, S. and Laeven, L. (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 36(3):563–583.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

Dahl, M. S. and Sorenson, O. (2010). The migration of technical workers. *Journal of Urban Economics*, 67(1):33–45.

Davidson, R. and MacKinnon, J. G. (1995). Estimation and inference in econometrics. *Econometrica*, 63(5):217–224.

De, P., Singh, A. E., Wong, T., and Kaida, A. (2007). Predictors of gonorrhea reinfection in a cohort of sexually transmitted disease patients in Alberta, Canada, 1991–2003. *Sexually Transmitted Diseases*, 34(1):30–36.

de Figueiredo, R. J. and Edwards, G. (2007). Does private money buy public policy? Campaign contributions and regulatory outcomes in telecommunications. *Journal of Economics & Management Strategy*, 16(3):547–576.

Depken, C. A. (1998). The effects of campaign contribution sources on the congressional elections of 1996. *Economics Letters*, 58(2):211–215.

Doherty, I. A., Padian, N. S., Marlow, C., and Aral, S. O. (2005). Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *The Journal of Infectious Diseases*, 191(1):42–54.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150.

Eames, K. T. and Keeling, M. J. (2004). Monogamous networks and the spread of sexually transmitted diseases. *Mathematical Biosciences*, 189(2):115–130.

Eichman, H., Hunt, G. L., Kerkvliet, J., and Plantinga, A. J. (2010). Local employment growth, migration, and public land policy: Evidence from the Northwest Forest Plan. *Journal of Agricultural and Resource Economics*, pages 316–333.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspective*, 3(4):99–117.

Esty, D. C. and Caves, R. E. (1983). Market structure and political influence: New data on political expenditures, activity, and success. *Economic Inquiry*, 21(1):24–38.

Faccio, M. (2006). Politically connected firms. *American Economic Review*, 96(1):369–86.

Federal Election Commission (2014). The FEC and the Federal Campaign Finance Law. `https://www.fec.gov/pages/brochures/fec_feca_brochure.pdf`. Accessed June 21, 2017.

Fenton, K. A., Johnson, A. M., McManus, S., and Erens, B. (2001). Measuring sexual behaviour: Methodological challenges in survey research. *Sexually Transmitted Infections*, 77(2):84–92.

Fernandez de Guevara, J., Maudos, J., and Perez, F. (2005). Market power in European banking sectors. *Journal of Financial Services Research*, 27(2):109–137.

File, T. (2015). Who votes? Congressional elections and the American electorate: 1978–2014. `https://www.census.gov/content/dam/Census/library/publications/2015/demo/p20-577.pdf`. Accessed June 21, 2017.

Fisman, R. (2001). Estimating the value of political connections. *American Economic Review*, 91(4):1095–1102.

Flom, P. L., Friedman, S. R., Kottiri, B. J., Neaigus, A., Curtis, R., Des Jarlais, D. C., Sandoval, M., and Zenilman, J. M. (2001). Stigmatized drug use, sexual partner concurrency, and other sex risk network and behavior characteristics of 18-to 24-year-old youth in a high-risk neighborhood. *Sexually Transmitted Diseases*, 28(10):598–607.

Goldberg, P. K. and Maggi, G. (1999). Protection for sale: An empirical investigation. *American Economic Review*, 89(5):1135–1155.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.

Gordon, S. C., Hafer, C., and Landa, D. (2007). Consumption or investment? On motivations for political giving. *Journal of Politics*, 69(4):1057–1072.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Grossman, G. M. and Helpman, E. (1994). Protection for sale. *American Economic Reveiew*, 84(2):833–50.

Grossman, G. M. and Helpman, E. (1996). Electoral competition and special interest politics. *The Review of Economic Studies*, 63(2):265–286.

Grossman, G. M. and Helpman, E. (2001). *Special interest politics*. MIT press.

Hainmueller, J. and Kern, H. L. (2008). Incumbency as a source of spillover effects in mixed electoral systems: Evidence from a regression-discontinuity design. *Electoral Studies*, 27(2):213–227.

Halperin, D. T. and Epstein, H. (2004). Concurrent sexual partnerships help to explain Africa's high HIV prevalence: Implications for prevention. *The Lancet*, 364(9428):4–6.

Hansen, W. L., Mitchell, N. J., and Drope, J. M. (2005). The logic of private and collective action. *American Journal of Political Science*, 49(1):150–167.

Harrington, K. F., DiClemente, R. J., Wingood, G. M., Crosby, R. A., Person, S., Oh, M. K., and HOOK III, E. W. (2001). Validity of self-reported sexually transmitted diseases among african american female adolescents participating in an HIV/STD prevention intervention trial. *Sexually Transmitted Diseases*, 28(8):468–471.

Hartman, E., Grieve, R., Ramsahai, R., and Sekhon, J. S. (2015). From sample average treatment effect to population average treatment effect on the treated: Combining experimental with observational studies to estimate population treatment effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3):757–778.

Hogben, M. and Leichliter, J. S. (2008). Social determinants and sexually transmitted disease disparities. *Sexually Transmitted Diseases*, 35(12):13–18.

Hser, Y.-I., Maglione, M., and Boyle, K. (1999). Validity of self-report of drug use among STD patients, ER patients, and arrestees. *The American Journal of Drug and Alcohol Abuse*, 25(1):81–91.

Hsieh, C.-S. and Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.

Igan, D. and Mishra, P. (2011). Three's company: Wall Street, Capitol Hill, and K Street. Technical report, International Monetary Fund.

Igan, D., Mishra, P., and Tressel, T. (2011). A fistful of dollars: Lobbying and the financial crisis. Technical report, IMF Working Papers 09/287.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jiménez, G., Lopez, J. A., and Saurina, J. (2013). How does competition affect bank risk-taking? *Journal of Financial Stability*, 9(2):185–195.

Johnes, G. and Hyclak, T. (1994). House prices, migration, and regional labor markets. *Journal of Housing Economics*, 3(4):312–329.

Johnston, R. J. (2014). *Money and votes (Routledge library editions: Political geography): Constituency campaign spending and election results*, volume 4. Routledge.

Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1–8.

Kelly, C. A., Soler-Hampejsek, E., Mensch, B. S., and Hewett, P. C. (2013). Social desirability bias in sexual behaviour reporting: Evidence from an interview mode experiment in rural Malawi. *International Perspectives on Sexual and Reproductive Health*, 39(1):14.

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251.

Kim, J.-H. (2008). Corporate lobbying revisited. *Business and Politics*, 10(2):1–23.

Kingston-Riechers, J. (2011). The economic cost of HIV/AIDS in Canada. `http://www.cdnaids.ca/the-economic-cost-of-hivaids-in-canada/`.

Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society*, 4(2):133–156.

Koetter, M., Kolari, J. W., and Spierdijk, L. (2012). Enjoying the quiet life under deregulation? Evidence from adjusted lerner indices for U.S. banks. *Review of Economics and Statistics*, 94(2):462–480.

Kossinets, G. and Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450.

Kroszner, R. S. and Strahan, P. E. (1999). What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *Quarterly Journal of Economics*, 114(4):1437–1467.

Kroszner, R. S. and Stratmann, T. (1998). Interest-group competition and the organization of Congress: Theory and evidence from financial services' political action committees. *American Economic Review*, 88(5):1163–1187.

Kustec, T., Keše, D., and Klavs, I. (2016). Under-reporting of sexually transmitted infection with chlamydia trachomatis- A revision of surveillance system is required. *Slovenian Journal of Public Health*, 55(3):174–178.

Laumann, E. O. (1994). *The social organization of sexuality: Sexual practices in the United States*. University of Chicago Press.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Liljeros, F., Edling, C. R., and Amaral, L. A. N. (2003). Sexual networks: Implications for the transmission of sexually transmitted infections. *Microbes and Infection*, 5(2):189–196.

Lin, H., He, N., Zhou, S., Ding, Y., Qiu, D., Zhang, T., and Wong, F. Y. (2013). Behavioral and molecular tracing of risky sexual contacts in a sample of Chinese HIV-infected men who have sex with men. *American Journal of Epidemiology*, 177(4):343–350.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224.

Macher, J. T., Mayo, J. W., and Schiffer, M. (2011). The influence of firms on government. *The B.E. Journal of Economic Analysis and Policy*, 11(1):1–27.

Martin, S. (1993). *Advanced industrial economics.* Blackwell.

Maudos, J. and Nagore, A. (2005). *Explaining market power differences in banking: A cross-country study.* Instituto Valenciano de Investigaciones Económicas.

Mayda, A. M., Ludema, R. D., and Mishra, P. (2010). Protection for free? The political economy of U.S. tariff suspensions. Technical report, International Monetary Fund.

Mayer, K. H. and Beyrer, C. (2007). HIV epidemiology update and transmission factors: risks and risk contexts?16th international AIDS conference epidemiology plenary. *Clinical Infectious Diseases*, 44(7):981–987.

McPherson, M., Smith-Lovin, L., and Brashears, M. E. (2006). Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review*, 71(3):353–375.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.

Mian, A., Sufi, A., and Trebbi, F. (2010). The political economy of the U.S. mortgage default crisis. *American Economic Review*, 100(5):1967–98.

Mian, A. R. and Khwaja, A. I. (2005). Do lenders favor politically connected firms? Rent provision in an emerging financial market. *Quarterly Journal of Economics*, 120(4):1371–1411.

Mitra, D., Thomakos, D. D., and Ulubaşoğlu, M. A. (2002). Protection for sale? In a developing country: Democracy vs. dictatorship. *Review of Economics and Statistics*, 84(3):497–508.

Molyneux, P., Lloyd-Williams, D. M., and Thornton, J. (1994). Competitive conditions in European banking. *Journal of Banking & Finance*, 18(3):445–459.

Morris, M. and Kretzschmar, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17(3-4):299–318.

Morton, R. B. (1987). A group majority voting model of public good provision. *Social Choice and Welfare*, 4(2):117–131.

Morton, R. B. (1991). Groups in rational turnout models. *American Journal of Political Science*, 35(3):758–776.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of group.* Harvard University Press Cambridge.

Olson, M. (2009). *The logic of collective action*, volume 124. Harvard University Press.

Panzar, J. C. and Rosse, J. N. (1987). Testing for monopoly equilibrium. *Journal of Industrial Economics*, 35(4):443–456.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979.

Perez Saiz, H. and Semenov, A. (2015). The effect of campaign contributions on state banking regulation and bank expansion in the U.S. Technical report, University of Ottawa.

Pittman, R. (1977). Market structure and campaign contributions. *Public Choice*, 31(1):37–52.

Plantinga, A. J., Détang-Dessendre, C., Hunt, G. L., and Piguet, V. (2013). Housing prices and inter-urban migration. *Regional Science and Urban Economics*, 43(2):296–306.

Pongou, R. (2009). Anonymity and infidelity: Ethnic identity, strategic cross-ethnic sexual network formation, and HIV/AIDS in Africa. Technical report, Department of Economics, University of Ottawa.

Pongou, R. and Serrano, R. (2009). A dynamic theory of fidelity networks with an application to the spread of HIV/AIDS. Technical report, Working Paper, Brown University, Department of Economics.

Pongou, R. and Serrano, R. (2013). Fidelity networks and long-run trends in HIV/AIDS gender gaps. *The American Economic Review*, 103(3):298–302.

Posner, D. N. (2004). Measuring ethnic fractionalization in Africa. *American Journal of Political Science*, 48(4):849–863.

Potterat, J. J., Muth, S., Rothenberg, R., Zimmerman-Rogers, H., Green, D., Taylor, J., Bonney, M., and White, H. (2002). Sexual network structure as an indicator of epidemic phase. *Sexually Transmitted Infections*, 78(1):152–158.

Public Health Agency of Canada (2015). Report on sexually transmitted infections in Canada: 2012. `http://www.catie.ca/sites/default/files/Report-on-STIs-in-Canada-2012.pdf`.

Rajan, R. G. and Zingales, L. (2003). The great reversals: The politics of financial development in the twentieth century. *Journal of Financial Economics*, 69(1):5–50.

Riker, W. H. and Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review*, 62(1):25–42.

Salamon, L. M. and Siegfried, J. J. (1977). Economic power and political influence: The impact of industry structure on public policy. *American Political Science Review*, 71(3):1026–1043.

Schaeck, K. and Cihak, M. (2012). Banking competition and capital ratios. *European Financial Management*, 18(5):836–866.

Schaeck, K., Cihak, M., and Wolfe, S. (2009). Are competitive banking systems more stable? *Journal of Money, Credit and Banking*, 41(4):711–734.

Sen, A. and Luong, M. (2008). Estimating the impact of beer prices on the incidence of sexually transmitted diseases: Cross-province and time series evidence from Canada. *Contemporary Economic Policy*, 26(4):505–517.

Shachar, R. and Nalebuff, B. (1999). Follow the leader: Theory and evidence on political participation. *American Economic Review*, 89(3):525–547.

Shaffer, S. (1993). A test of competition in Canadian banking. *Journal of Money, Credit and Banking*, 25(1):49–61.

Snyder, J. M. (1990). Campaign contributions as investments: The U.S. House of Representatives, 1980-1986. *Journal of Political Economy*, 98(6):1195–1227.

Snyder, J. M. (1993). The market for campaign contributions: Evidence for the U.S. Senate 1980–1986. *Economics and Politics*, 5(3):219–240.

Strassmann, W. P. (2000). Mobility and affordability in U.S. housing. *Urban Studies*, 37(1):113–126.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64(5):416–424.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Treas, J. and Giesen, D. (2000). Sexual infidelity among married and cohabiting Americans. *Journal of marriage and family*, 62(1):48–60.

Tuite, A. R., Jayaraman, G. C., Allen, V. G., and Fisman, D. N. (2012). Estimation of the burden of disease and costs of genital chlamydia trachomatis infection in Canada. *Sexually Transmitted Diseases*, 39(4):260–267.

Tullock, G. (1972). The purchase of politicians. *Western Economic Journal*, 10(3):354–55.

Uhlaner, C. J. (1989). Rational turnout: The neglected role of groups. *American Journal of Political Science*, 33(2):390–422.

Vives, X. (2011). Competition policy in banking. *Oxford Review of Economic Policy*, 27(3):479–497.

Wasserheit, J. N. (1992). Epidemiological synergy: Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. *Sexually Transmitted Diseases*, 19(2):61–77.

Wiesenfeld, H. C., Hillier, S. L., Meyn, L. A., Amortegui, A. J., and Sweet, R. L. (2012). Subclinical pelvic inflammatory disease and infertility. *Obstetrics and Gynecology*, 120(1):37–43.

Wylie, J. L. and Jolly, A. (2001). Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. *Sexually Transmitted Diseases*, 28(1):14–24.

Zuma, K., Gouws, E., Williams, B., and Lurie, M. (2003). Risk factors for HIV infection among women in Carletonville, South Africa: Migration, demography and sexually transmitted diseases. *International Journal of STD and AIDS*, 14(12):814–817.